ARTICULATION CORRECTION OF THE DEAF BY MEANS OF VISUALLY DISPLAYED ACOUSTIC INFORMATION

D. J. L. POVEL

ARTICULATION CORRECTION OF THE DEAF BY MEANS OF VISUALLY DISPLAYED ACOUSTIC INFORMATION

Promotor: Prof. Dr. W.J.M. Levelt.

ARTICULATION CORRECTION OF THE DEAF BY MEANS OF VISUALLY DISPLAYED ACOUSTIC INFORMATION

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE SOCIALE WETENSCHAPPEN AAN DE KATHOLIEKE UNIVERSITEIT TE NIJMEGEN, OP GEZAG VAN DE RECTOR MAGNIFICUS PROF. Mr. F.J.F.M. DUYNSTEE VOLGENS HET BESLUIT VAN HET COLLEGE VAN DECANEN IN HET OPENBAAR TE VERDEDIGEN OP WOENSDAG 18 DECEMBER 1974 DES NAMIDDAGS TE 3.00 UUR PRECIES

DOOR

DIRK JAN LODEWIJK POVEL

GEBOREN TE BREDA

Uitvoering en druk. Th.J.M. Fuchten en J.P.H. Benchop

Naast de personen die reeds op andere plaatsen in dit proefschrift zijn vermeld, wil ik hier nog graag mijn dank betuigen aan drs. Arno Hanraets, drs. Anneke Kremers, drs. Hanneke Receveur, drs. Clemens Janzing en Tiny Goessens voor hun hulp bij het opnemen, analyseren, afbeelden, beluisteren en identificeren van enkele duizenden klinkers.

CONTENTS

PREFACE

This study originates from an interest in the possible applications of 'visible speech apparatus' (i.e. devices that visually display aspects of the acoustic speech signal) for speech correction of the deaf.

Most devices that have been constructed to display one or more acoustic speech parameters (intonation, intensity, spectral composition, temporal structure) stem from a technical engineering background and were not directed towards a practical application such as for instance speech training of the deaf.

Potter, Kopp and Green (1947) were the first to apply 'visible speech' as an aid for the deaf, albeit with regard to speech perception rather than speech correction. This example was followed by several attempts to use newly developed visible speech apparatus for the benefit of speech correction of the deaf.

Most of these attempts broke down in their earliest developmental stages and did not result in the construction of a practicable speech correction device, either because of an inadequate definition of the requirements for a speech corrector or because of as yet unsolved technical problems in the construction of the apparatus. For these reasons we do find many indications of potential speech correction devices in the literature, but very little research devoted to the actual applicability of visible speech apparatus.

The present study starts with the formulation of a theoretical framework in terms of which a number of relevant aspects of speech production and speech acquisition both of the hearing and of the deaf child are described. In particular, it offers an opportunity for appreciating the implications of the alternative method of speech correction which consists in providing information about the speech product.

Moreover, from the theoretical descriptions, a number of requirements are inferred that serve as a guide in the construction of the Vowel Corrector, a device that can be described as an articulation corrector which displays spoken vowels as points in a two dimensional space.

This Vowel Corrector finally is evaluated with respect to its practical feasibility as an aid in the articulation correction of the deaf.

1. SPEECH PRODUCTION

1.1. Introduction

Under normal circumstances the acquisition of speech of the young child seems to develop of its own accord. In contrast, for instance, with learning to write, the educators of the child do not have to be actively engaged in this learning process. It is true that the adult, when speaking to the young child, will employ simple sentences, clearly mark sentence boundaries, and speak with an emphatic intonation (ref. Levelt 1975), but his real activity is restricted to the occasional correction of wrong grammar, incorrect use of words or speech errors. When the natural process of development is severely interfered with, as in the case of a deaf, the above no longer applies. On the contrary, whether the deaf child learns to speak will depend principally on the active speech instruction given by his teachers.

In practice, the speech acquisition of the deaf will form an integrated part of his language education (for the problem of language teaching of the deaf as a whole, see Van Uden 1968, 1974). However, we wish to make a sharp distinction between aspects of language and of speech, and we shall here be concerned only with the speech aspect or, more specifically, with the aspect of correct pronunciation. With regard to its production technique, correct pronunciation is dependent on correct respiration (c.f. Lenneberg 1967, Stetson 1951), on correct phonation (c.f. Bouhuys 1968) and on correct articulation. The present study is confined to the articulatory aspect of speech production since the visible speech apparatus that has been developed and that will be described below, displays only articulatory features.

An important part of our study is devoted to determining the theoretical significance of displaying visual aspects of the speech signal in speech training of the deaf. In this context, a distinction will be made with respect to the types of information that can be supplied to the deaf subject during acquisition. The information supplied can be related to motor aspects of speaking (e.g. the articulatory configurations used in speaking) or to acoustic aspects of the speech product itself. This dis-

tinction is elaborated in Chapter 2.

The significance of acoustic information in the process of speech acquisition can only be evaluated adequately when a proper insight has been obtained into the relation between the acoustic correlate and articulatory behaviour. Therefore, the present chapter will study the process of speech production in considerable detail.

In order to obtain a better view of the complicated mechanisms involved in the process of speaking, an analogy from the field of skilled performances has been chosen and worked out. This analogy is described in Section 1.2. Next, Section 1.3.1 deals with the problem of the unit of speech production.

1.2. Speech production analogy with flute playing

In order to elucidate the many complicated problems connected with the process of speaking, we have looked for a suitable analogy, an analogy enabling us to see this process objectively and clearly. The flute-playing analogy seemed to serve the purpose very well.

At first sight a certain similarity between the vocal tract and the flute seems apparent. As with the mouth, so with the flute, different sounds are formed by means of modifying the resonant characteristics of a cavity. The flute-player achieves these modulations by opening and closing the holes of his flute, the speaker by changing the adjustment of his articulators. Below, we will examine in detail the flute-playing process.

Flute playing

Playing the flute can be defined as the production of tones in the temporal domain. Here we will in first instance set aside the temporal aspect and concentrate on the production only of the tones. On the elaboration of the analogy we have in mind a very simple flute, the recorder, which possesses 8 holes that can be closed with the fingers. Various tones can be generated by forming different patterns of closed and opened holes.

People who start playing the flute will mostly assume that the important thing to be 'learnt' consists in the knowledge of the different patterns of opened and closed holes (finger positions), because these finger positions will enable the player to produce the tones which make up the language of music. The total number of tones that can be produced on the recorder (restricting ourselves to West-European music) amounts to approximately 24. The number of finger positions exceeds this number, since for most tones there are alternative finger positions to facilitate the playing of certain sequences. Thus the total number of finger positions amounts to approximately 40. The aspirant flute player will soon find out that the knowledge of these finger positions, however essential, is insufficient when actually playing. He will also have to learn how the desired finger positions must be realized from all possible

initial positions of the fingers. Then he will discover that, depending on the actual starting point, the realization of a finger position may in one instance be very easy or, in another, most difficult. This will be illustrated in two examples. (Figure 1 and 2)

Finger position for tone	A	B	C
	1 X	1 X	1 O
	2 X	2 O	2 X
	3 O	3 X	3 O
	4 O	4 X	4 O
	5 O	5 O	5 O
	6 O	6 O	6 O
	7 O	7 O	7 O

Figure 1. Finger positions for the production of three tones. (The letters do not correspond to names of notes)
X = closed hole;
O = opened hole.

For convenience we shall indicate the fingers by the digits 1 to 7. Since, in the examples given, the eighth hole does not play a role, it will not be mentioned. In order to produce tone C starting from either position A or position B, the following, very distinct actions will have to be performed:

A → C : raise finger 1

B → C : raise fingers 1, 3 and 4;
press down finger 2

We hereby assume that fingers which do not change position do not receive a command.
In order, in our second example (Figure 2, p. 6), to realize tone E starting from either D or F, the following actions will be required:

D → E : raise finger 6

F → E : press down finger 5

Our first example clearly demonstrates that the complexity of the actions for the realization of one finger position may vary considerably depend-

Finger position for tone	D	F	E
	1 X	1 X	1 X
	2 X	2 X	2 X
	3 X	3 X	3 X
	4 X	4 X	4 X
	5 X	5 O	5 X
	6 X	6 O	6 O
	7 O	7 O	7 O

Figure 2. Finger positions for the production of three tones.

ing on the point of departure. The second example shows that there is not necessarily a common aspect in the various actions that may serve to realize one particular finger position.

It will now be clear that the flute player not only needs to know the finger positions of the different notes, but he must also have at his disposal a repertoire of actions which enable him to produce those finger positions whatever his initial position is. If we assume that all sequences of two tones can occur (which seems reasonable) then the repertoire will consist of $P(^{24}_{2}) = 552$ actions. This calculation does not take into account the actions required for the production of the auxiliary finger positions. If these actions are included, the total number will be between 1000 and 1500.

The most salient feature of the auxiliary finger positions is, that they facilitate the transitional movement from one tone to the next. This will be of great importance in fast passages. In Figure 3, the difference between a standard and an auxiliary finger position is shown.

Finger position for tone	A	A^1	B
	1 X	1 X	1 O
	2 O	2 X	2 X
	3 X	3 O	3 O
	4 X	4 O	4 O
	5 O	5 O	5 O
	6 X	6 O	6 O
	7 O	7 O	7 O

Figure 3. Alternative finger positions for the production of tone A, and a finger position for tone B.

A^1 is an auxiliary finger position for A. The required actions to perform B starting from A or A^1 are, respectively,

$A \rightarrow B$ raise fingers 1, 3, 4 and 6
press down finger 2

$A^1 \rightarrow B$ raise finger 1

It should be noted that the choice of an auxiliary finger position for a particular tone can be determined by the characteristics of the preceeding tone as well as of the succeeding tone. Here, conflicts can arise. Take for instance the case in which a tone sequence A B C has to be produced, where B can be realized by means of different finger positions. It might be the case that, following A, an auxiliary finger position for B is possible or even desirable, but that this finger position forms an unfavourable starting position for the production of tone C. We should finally remark that a finger position will be termed auxiliary if the tone thus produced is of a poorer quality than the tone produced with the standard finger position.

In the simplest case the input for the flute player consists in a discrete code in which the tones to be produced and their temporal structure are specified. The process of playing the flute can now be described as follows. In a first stage the player determines which finger position will be used to produce each element in the code (note). If there are different finger positions available for one note the player must make a choice, taking into account the context and the duration of the element in question. Thus the original sequence of notes is transformed into a sequence of finger positions. By paired comparison of the elements of the latter, second code, the player decides in a second stage upon the actions needed for the realization of the successive finger positions. This latter sequence, finally, still has to be translated into commands to the proper muscles. The process thus far can be represented in a flow diagram. (Fig. 4)

Notes	Transformation	Transformation	Transformation	Muscles
→	into	→ into	→ into	→
(In)	finger positions	actions	neural commands	(Out)

Figure 4. Flow diagram for the transformation of notes into neural commands to the muscles.

Finger position for tone	D	F	E
	1 X	1 X	1 X
	2 X	2 X	2 X
	3 X	3 X	3 X
	4 X	4 X	4 X
	5 X	5 O	5 X
	6 X	6 O	6 O
	7 O	7 O	7 O

Figure 2. Finger positions for the production of three tones.

ing on the point of departure. The second example shows that there is not necessarily a common aspect in the various actions that may serve to realize one particular finger position.

It will now be clear that the flute player not only needs to know the finger positions of the different notes, but he must also have at his disposal a repertoire of actions which enable him to produce those finger positions whatever his initial position is. If we assume that all sequences of two tones can occur (which seems reasonable) then the repertoire will consist of $P(\binom{24}{2}) = 552$ actions. This calculation does not take into account the actions required for the production of the auxiliary finger positions. If these actions are included, the total number will be between 1000 and 1500.

The most salient feature of the auxiliary finger positions is, that they facilitate the transitional movement from one tone to the next. This will be of great importance in fast passages. In Figure 3, the difference between a standard and an auxiliary finger position is shown.

Finger position for tone	A	A^1	B
	1 X	1 X	1 O
	2 O	2 X	2 X
	3 X	3 O	3 O
	4 X	4 O	4 O
	5 O	5 O	5 O
	6 X	6 O	6 O
	7 O	7 O	7 O

Figure 3. Alternative finger positions for the production of tone A, and a finger position for tone B.

A^1 is an auxiliary finger position for A. The required actions to perform B starting from A or A^1 are, respectively,

$A \rightarrow B$ raise fingers 1, 3, 4 and 6
press down finger 2

$A^1 \rightarrow B$ raise finger 1

It should be noted that the choice of an auxiliary finger position for a particular tone can be determined by the characteristics of the preceeding tone as well as of the succeeding tone. Here, conflicts can arise. Take for instance the case in which a tone sequence A B C has to be produced, where B can be realized by means of different finger positions. It might be the case that, following A, an auxiliary finger position for B is possible or even desirable, but that this finger position forms an unfavourable starting position for the production of tone C. We should finally remark that a finger position will be termed auxiliary if the tone thus produced is of a poorer quality than the tone produced with the standard finger position.

In the simplest case the input for the flute player consists in a discrete code in which the tones to be produced and their temporal structure are specified. The process of playing the flute can now be described as follows. In a first stage the player determines which finger position will be used to produce each element in the code (note). If there are different finger positions available for one note the player must make a choice, taking into account the context and the duration of the element in question. Thus the original sequence of notes is transformed into a sequence of finger positions. By paired comparison of the elements of the latter, second code, the player decides in a second stage upon the actions needed for the realization of the successive finger positions. This latter sequence, finally, still has to be translated into commands to the proper muscles. The process thus far can be represented in a flow diagram. (Fig. 4)

Notes	Transformation	Transformation	Transformation	Muscles
→	into	→ into	→ into	→
(In)	finger positions	actions	neural commands	(Out)

Figure 4. Flow diagram for the transformation of notes into neural commands to the muscles.

From the flow diagram given one could conclude that the process under consideration is an open loop process, i.e. does not make use of feedback.

Since the problem of the deaf, as regards his attempts to speak, is caused by feedback deprivation, we shall also in the flute comparison look in considerable detail at the role of feedback.

The role of feedback in flute playing

The sensory feedback that the subject receives during (and after) the performance of skilled behaviour can, with respect to its function in the process, be divided into two types. The first type of feedback is information that is used for the computation of the on-going behaviour, the second type of feedback has no role in the control of the behaviour but fulfills a 'monitoring' function. It does not seem likely that all the feedback that is available during flute playing will be of the first type, since this process consists of achieving continually changing targets (the finger positions). Whether feedback can be used for the computation of the on-going behaviour is not only dependent on the frequency of changes of target but also on the duration of the feedback loops.

Feedback of the second type is information that arrives centrally too late to be used for the control of the on-going behaviour. This information can, however, be used as a check on already generated behaviour and when a discrepancy is detected between that behaviour and a certain criterion it will be stopped and may be corrected This type of information presumably plays a major role in the acquisition stage.

We will now give a more detailed description of that part of the process of flute playing in which feedback may be expected to play a role. This is the stage of the process in which the sequence of finger positions is transformed into a sequence of actions. Theoretically there are two possible alternatives which will be treated successively.

1. Open loop. In the open loop model, every action will be determined centrally, it will not make use of feedback on the actual state of the peripheral organs. The knowledge about this state, necessary in determining the actions to follow, may be assumed to be centrally available, since this state was the target of the preceeding action. This type of behaviour regulation is described by Keele (1968) as controlling by means of a motor program. According to Keele, 'the concept of a motor

program may be viewed as a set of muscle commands that are structured before a movement begins and that allows the entire sequence to be carried out uninfluenced by peripheral feedback.' (p. 387). Within a model of this type, errors may arise when for some reason the state in the periphery differs from the state that is centrally expected to exist. Such a situation may give rise to an accumulation of errors.

2. Closed loop. In the closed loop model, information concerning the state of the peripheral organs is not assumed to be centrally available. It is thought that information is continuously incoming from the periphery via afferent channels. The flow diagram for this model is presented in Figure 5.

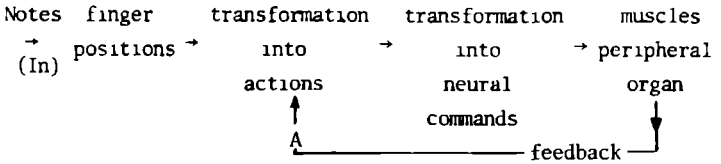

Figure 5. Flow diagram illustrating the transformation of notes into neural commands for the muscles. Closed loop

A model such as pictured in the flow diagram above will prove particularly useful in those cases where there is central uncertainty regarding the peripheral state. This may be the case if neural commands do not result in the intended state of the peripheral organs.
The latter model will only work satisfactorily however if the feedback loop (A) is short compared with the interval time between successive states. The fact is that a new action can only be computed once the feedback on the existing peripheral state – brought about by the preceeding action – has arrived centrally (see Figure 6).

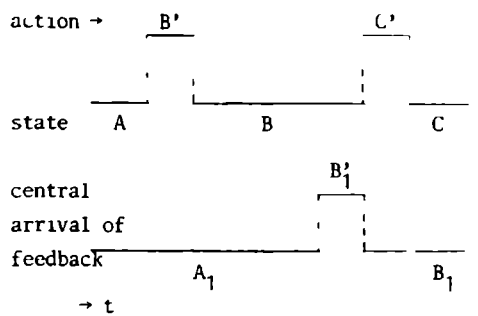

Figure 6. Peripheral state and central arrival of feedback as a function

of time (t).

Upper part:

A, B, C are states (finger positions)

B' is the action that changes A → B

C' is the action that changes B → C

lower part

A_1 *is feedback from state A*

B'_1 *is feedback from action B'*

B_1 *is feedback from state B*

Woodworth (1899) already noticed that rapid, successive movements cannot be controlled by means of visual feedback. Lashley (1951) rules out all forms of sensory control of quick successions of movements occurring in piano playing. Even the duration of the relatively fast proprioceptive feedback is still long compared with the interval time in fast tone sequences. According to Craik (1947), 150 ms. pass before a reaction to a proprioceptive change in arm or hand can be noticed. Gibbs (1965) finds in a quite different setting 110 ms., while Lashley (1951) reports 125 ms. These durations rule out the possibility of proprioceptive control of movements in fast tone sequences.

There are instances, however, where the flute player necessarily has to rely on different information. This will be the case when it is centrally not known what the peripheral state looks like. This may occur when an error has been made in the determination of actions required for the attainment of a specific state. The feedback used in such cases can be visual or tactile-kinesthetic (possibly auditory). It can also frequently be observed that when a flute player starts to play, he takes notice of the position of his fingers with respect to the holes of the flute, either visually or tactually.

Summarizing, we can characterize the process of flute playing as transformations of the input, which normally is made up of a code of tones. In a first transformation, the notes are translated into finger positions. Next, the sequence of finger positions is translated into a sequence of actions. Feedback possibly plays a role in the determination of the actions, this is, however, not necessarily so and would even seem very unlikely in very fast tone sequences. In this context the grouping assumption made in the literature on skilled performance should be mentioned. This assumption is that what was first a sequence of independent units becomes organized as one new programmed unit. Besides a redefinition of the concept of the unit of production (actually an al-

most infinite extension of the number of units), it implies that within chunks no use of feedback is made.

In addition to its function in the control of behaviour, feedback, whether it be auditory, tactile, kinesthetic or visual, can be employed in the continuous monitoring system of on-going behaviour. In this system, feedback from the different sensory modes is matched with a norm that defines what the different feedback aspects should look like at each moment. If a mistake has been made, the sequence containing an error may be repeated.

We shall now turn to speech production and examine to what extent the flute analogy can be a model for the process of speaking.

1.3. Speaking

Although there is some apparent resemblance between the flute and the mouth as instrument for the production of sounds on the one hand, and between flute playing and speaking on the other, we shall now consider a number of differences. The flute has been designed for the production of a restricted set of sounds that form the alphabet of a (musical) language, whereas the mouth can produce not only a gamut of non-speech sounds but, in principle, all sounds of all languages in the world. Moreover, the flute has only a limited number of discrete states, the mouth, in contrast, can be in an infinite number of states, modulating continuously from one state into the next. The distinctions mentioned will presumably have important implications for the learning processes that underlie flute playing or speaking. This subject will be treated in Chapter 2. The distinction between discrete and continuous made above has its counterpart in the field of perception and is manifested in the role of the transitional phenomena. In sound sequences produced on the flute, the transitions are steplike, they are of very short duration and, in principle, alike for all transitional movements. In sequences of speech sounds on the other hand, the transitions are never steplike, have a relatively long duration and differ, depending on what transition is realized. Indeed, they appear to form a highly informational part of the speech signal and therefore to be of great importance to speech perception. (Liberman 1957, Liberman et al. 1967)

With respect also to the processes that underlie flute playing and speaking we can point out a number of differences. In the process of flute playing, for instance, it appears very well possible to indicate the units, in the process of speaking, however, the definition of a unit of production will appear to be much more difficult. This aspect will be discussed below. Another difference lies in the specifiability of the actions required for flute playing and for speaking. Though to some extent we are able to specify the necessary behaviour for playing the flute, this will appear to be a great problem for speech production. This subject will also be treated in some detail below.

When the number of sounds produced per unit of time in flute play-

ing and speaking is compared, it can be observed that the maximum rate of speaking is higher than the maximum rate of producing tones on the flute. Maxima of tone production in very fast passages lie between 10 and 16 tones per second. The maximum speed of producing speech sounds is estimated to vary from 18 to 24 sounds (phonemes) per second (Lenneberg 1967). Such rates can be maintained for short periods of time only (Goldman-Lisler 1954). For the effect of this rate of speaking on perception see Liberman et al. (1967).

Finally, we may expect to find differences between the possible role of feedback in both processes, because of the fact that the length of the feedback loops is shorter in speaking than in flute playing. The implications of acoustic, tactile and kinesthetic feedback in the process of speech production will be given considerable attention below.

Though we are able to point out differences between flute playing and speaking, yet as a starting point for our presentation it seems justifiable to consider the mouth as an instrument that can be used for the generation of sound sequences. Below we will examine to what extent the analogy is applicable to the production of speech sounds.

1.3.1. Units

In first instance, the evidence concerning the presence of units in speech comes from perception. In the speech signal, sentences, words, syllables and sounds can readily be identified. For our purpose, however, we are not really interested in the existence of perceptual units and therefore we do not wish to consider the pertaining research but rather to trace the identity of the production units in speaking.

The most important evidence with respect to the subject comes from speech errors, to which a considerable number of studies are devoted. (See list of references from studies cited below) Two thorough investigations of speech errors have been made by Nooteboom (1969) and Fromkin (1971). Fromkin, in the paper mentioned, has made some very interesting remarks concerning the theoretical relevance of the unit concept: 'What is apparent in the analysis and conclusions of all linguists and psychologists dealing with errors in speech is that, despite the semi-continuous nature of the speech signal, there are discrete units at some level of PERFORMANCE which can be substituted, omitted, transformed or added. It should be stated here that, were we to find no evidence in actual speech production or perception for such discrete units, this would not be sufficient cause to eliminate discrete units in phonology or syntax. The fact that it is impossible to describe the grammars of languages without such units is itself grounds for postulating them in a theory of grammar. But when one finds it similarly impossible to explain speech production (which must include errors made) without discrete performance units, this is further substantiation of the psychological reality of such discrete units.'

From the various studies it can be inferred that several units play a role in speech production. The majority of speech errors, however, refer to separate speech sounds (phonemic errors). In Nooteboom's study approximately 80% of the speech errors are phonemic errors. Three types of errors can, in general, be distinguished, viz.: anticipations, for example: 'Warold Wilson' instead of 'Harold Wilson'; perseverations, for example: 'Harold Hilson' instead of 'Harold Wilson'; and transpositions, for example: 'Warold Hilson' instead of 'Harold Wilson'. Anticipation is the most usual type forming approximately 75% of

the cases, perseverations and transpositions account for 20% and 5% respectively. In phonemic errors, vowels as well as consonants can be involved. Often, consonant clusters function as a group, e.g.'coat thrutting' instead of 'throat cutting'. Fromkin, however, also gives examples in which only one element of a consonant cluster is involved, e.g.'blake fruid' instead of'brake fluid'. This finding indicates that consonant clusters should not be regarded as inseparable units, but as composed of smaller units. Affricates and diphthongs always act as groups in speech errors.

Syllables and words also sometimes act as units that are omitted, added or transposed. The frequency of occurrence is low, however, in Nooteboom's collection of speech errors 9% and 2%, respectively. Although it is observed that clusters, syllables and words sometimes act as a whole, they should not be considered as indivisible, but instead as being made up of smaller elements – the sounds – that function more or less independently within the mentioned groups.

The independence of these elements is not complete: the slips of the tongue appear to follow certain rules. Thus, a prevocal or postvocal consonant will only be influenced by a prevocal or postvocal consonant, respectively, of an other syllable, the syllabic nucleus will only be influenced by the nucleus of an other syllable. Interactions giving rise to speech errors occur more frequently between similar elements, or between elements within similar phonetic contexts. Moreover, it appears that all occurring speech errors obey the phonological rules of language; in other words, sounds or sound sequences not occurring in the language will not appear in speech errors. This peculiar property is characteristic of speech errors: the freedom of movement of elements on a lower level are determined by construction rules of a higher level. Because of this property Fromkin stresses the hierarchical character of the process of speech production.

The 'distinctive features' form a special case. Nooteboom reports not to have found evidence which indicates that distinctive features behave as independent units feature substitutions do not occur in his collection. Fromkin, however, gives examples of speech errors which can be most easily explained by assuming that distinctive features behave as independent units. This holds for about 5% of her collection of speech errors. An example of feature transposition in which voicing is involved is: 'big and fat' → 'pig and vat'. b → p = +voice → -voice, f → v = -voice → +voice. She sums up: 'the only conclusion one can draw from the

examples of feature switching given above is that at least some of the proposed distinctive features are independent behavioral units.' (p. 37 o.c.). Since feature substitutions always result in an other speech sound, these mistakes are often classified as phonemic errors. For example, 'gall the curl' instead of 'call the girl' can be regarded as k → g or as -voice → +voice.

From the preceeding presentation, we infer two arguments to confirm the view that the phoneme is the building block in the process of speech production rather than the syllable. The syllable, notwithstanding its function as a structuring unit, cannot be considered to play this role In the first place, the fact that the number of syllables far exceeds the number of speech sounds (in English there are approximately 8000 syllables, ref. Heffner, 1969) forms a, be it indecisive, counter-argument. Secondly, and more importantly, is the following consideration. Let us suppose for the moment that the syllable is the building block in speech production. We would then expect those syllables that do not occur in a language not to appear in speech errors either, as is the case with phonemes. This, however, does appear to happen and we can cite two examples of the use of such non-existent syllables in Dutch, from an article by Cohen (1965), viz ·

'bruuk' in 'rubruuk' instead of 'rubriek'
'zuut' in 'U zuut' instead of 'U ziet'.

Though we realize that the reviewed data do not guarantee a definitive conclusion, we still support the phoneme as the most probable unit of speech production. Here one remark should be made as regards the use of the term phoneme. Perhaps it would be more correct to use the term 'phone' or 'speech sound' instead of 'phoneme' (as Fromkin (1971) does), as the concept 'phoneme' has certain theoretical connotations (e.g. bundle of distinctive features) not relevant in this connection. Since in most of the studies quoted below the term 'phoneme' is used and not 'phone', we shall adhere to the term phoneme.

1.3.2. Units in speech production

The phonemes, regarded as being the building blocks in the process of speaking, have the same function as the notes for the flute player. There is one important difference. The flute player has at his disposal a code in which the units are specified (the code of notes), the speaker, however, has not. He will either have to conceive the phonemes himself (in spontaneous speaking), or to infer them from an acoustic message that he wants to imitate, or, in the case of reading aloud, to transform the orthographic code into a sequence of phonemes. Below we shall examine the question as to whether the phonemes undergo similar transformations to those of the notes in the flute model or whether a one-to-one relationship exists between phonemes and neural commands to the articulators, as has been suggested in the literature.

So far, the phonemes have no objective status. They only exist in the head of the examiner who could infer them from speech errors. On the analogy of flute playing one would expect that these units can similarly be identified in the acoustic signal. It is well-known, however, that such 'phoneme invariance' is not found in the acoustic speech signal. Several authors have reported that, even in one speaker, the acoustic properties of different realizations of the same phoneme vary according to context, stress and rate of speaking. These phenomena, indicated by the terms coarticulation, reduction and neutralization, are described in detail by several authors. Liberman (1957, 1967) demonstrated the 'restructuring' of the phoneme in the acoustic signal, Lindblom (1963) showed in spectrographical analysis the considerable effects of stress and speaking rate, while Koopmans (1972) examined modifications of the phoneme as a function of speaking mode. The acoustical variation that is found is so large, even in singly spoken words, that it forms a real problem for procedures of automatic speech recognition which are based on the identification of separate phonemes.

If no invariance can be shown in the acoustic signal, one would expect to find no invariance either in the configuration of the speech apparatus during the production of the various variants of a phoneme (allophones) and, accordingly, no invariance in the neural commands to the speech apparatus. In the flute model we would explain the observed

differences in the acoustic signal by assuming that different articulatory configurations were used for the realization of a sound, depending on its context and the rate of production.

An influential group of investigators, the Haskins group, has developed a quite contrary hypothesis which states that to each phoneme there is one corresponding neural command. In this conception the observed acoustical variation is attributed to mechanical constraints of the articulators or to the effects of overlap of neural commands. In this context one should remember that the mouth is a continuous instrument. Therefore it is possible (in contrast to sounds produced on the flute) that a sound is perceptually acceptable even when the 'ideal' position of the articulators has not been realized. (The 'ideal' position is that position of the articulators that is reached and maintained for some time during the steady state of a slowly pronounced phoneme, with no context influences.) Also relevant to the hypothesis mentioned are the investigations by Liberman et al. (1967) which show that the information determining the identity of a phoneme in the acoustic signal is not confined to the corresponding segment only but is also to be found in adjacent segments.

A model which assumes a one-to-one relation between phoneme and neural command is attractive. Not only is it very simple, fitting the requirements of parsimony, it moreover provides the possibility of explaining the existence of the perceptual unit, viz., the phoneme. The motor theory of speech perception (Liberman et al. 1962) is, in fact, based on the assumption that, corresponding to the phonemes, there are invariant motor commands, on which the incoming acoustic signal is mapped. In the next paragraph the most rigid variant of such a model will be discussed. This model assumes invariant motor commands up to the level of muscle innervation corresponding to phonemes.

1.4.1. An alternative model: invariant motor commands

In his 1970 article, McNeilage reports 'In 1958 the Haskins group began an attempt to show that the EMG (electromyographic) correlates of the phoneme will prove to be invariant in some significant sense' (p. 183). This electromyographic research records muscle potentials by means of electrodes placed in or on the surface of muscles of the articulators. A description of the electromyographic technique is given by Fromkin and Ladefoged (1966) A direct relation is assumed between neural command and electromyogram.

It should be kept in mind that the flute model, even if there were only one position per phoneme, does not predict invariant EMGs. On the contrary, it supposes different actions for the realization of one particular position depending on the starting position. It seems reasonable to postulate that this process also applies to speaking. Indeed, the modifications of the articulators necessary for the attainment of a speech sound will depend quite as much on the initial position of the articulators. Consider for instance the difference of the movements that must be executed for the realization of the /p/ in the sequences /ap/ or /mp/.

The first EMG studies seemed to support the hypothesis of one invariant neural command to the articulatory muscles per phoneme, regardless of context, etc. However, closer investigations of these studies and later research furnished data that showed a considerable amount of variance in the EMGs corresponding to one phoneme.

Fromkin (1966), for instance, analysed electromyograms of the orbicularis orbis of a speaker that pronounced a list of monosyllables. The list was made up of CVCs of the types /dVd/, /bVd/, /dVb/, /bVb/ and /pVp/, in which the vowel V was systematically varied in all consonantal contexts. The most important measures used to determine the similarity between EMGs were the amplitude of the highest peak of the EMG signal, duration of muscle activity and the time interval between EMG trace and acoustic signal. The EMG signals corresponding to the articulations of an inital /b/were found to be relatively invariant, this was also the case for a final /b/, an initial /p/ and a final /p/. Considerable differences were observed, however, between EMG parameters of an init-

ial /b/ and a final /b/, as well as between an initial /p/ and a final /p/. As expected, the articulation of a /d/ did not produce any significant muscle activity. The /dVd/ series was only used to determine which vowels did demand activity by the orbicularis orbis This appeared to be the case for the rounded vowels. Fromkin could thus demonstrate that when a rounded vowel was preceeded by a bilabial consonant, the activity associated with the vowel was considerably reduced. Neither this latter finding, nor the mentioned differences that were found between initial and final /b/ and /p/, fit a model assuming invariant motor commands corresponding to phonemes.

McNeilage and De Clerk (1969) tested the same hypothesis in a study in which they made use of EMG recordings and of cinefluorograms. Cinefluorograms are obtained by filming the intra oral structures with a high speed camera after applying a film-sensitive material to the tongue, the lips and part of the palate (ref. Strenger 1968). EMG potentials were recorded by means of 8 electrodes placed on different articulators. In this study the subject also pronounced monosyllables, viz. all combinations of the consonants /b,d,g/ and the vowels /ɪ,u, æ,ɔ/. In all cases the investigators found coarticulation effects of preceeding phonemes in one or more aspects of the EMG, in most cases coarticulation effects of succeeding phonemes could be demonstrated as well. Effects of C_1 on C_2 (in C_1VC_2) were not found. The authors conclude that invariance down to the level of muscle contraction can certainly not be assumed.

A model based on invariant phoneme commands can always be maintained by assuming that invariance occurs at some higher level. For this reason McNeilage and De Clerk developed a model that indeed starts from invariant commands but that also possesses a number of mechanisms that may explain some peripheral variance. These mechanisms, namely, an 'Anticipatory' mechanism, a 'Compatibility' mechanism and a 'Gamma Loop' mechanism, partially based on physiological evidence, each predict a special type of coarticulation effects. The authors compared the predictions with the observed empirical coarticulation effects and concluded that not all effects can be explained by the hypothesized mechanisms. We shall later return to the most important of these mechanisms, the gamma-loop system.

In a recent study, Sussman et al. (1973) demonstrated not only interactions between neighbouring segments but also interactions between more distant segments. Thus a left-to-right effect of V_1 on V_2

in V_1CV_2 syllables was found. This finding will also prove to have consequences for a model which has yet to be discussed and which is based on context-dependent motor commands. For the moment, we conclude that phoneme-invariant commands on the level of actual muscle innervation are not found.

There are two possibilities of maintaining the principle of the model. One is to propose a different unit, e.g. allophones. Wickelgren (1969) has actually developed such a model that will be described below. A second possibility is to assume that the neural commands are put in terms of target positions for the articulators (go-to commands, ref. Liberman 1967), such that each phoneme corresponds to a target position. The model must incorporate a mechanism that copes with the actual realization of a target position. It should be noted that such a model does not predict a one-to-one relation between phoneme and motor command to the articulators (as inferred from EMGs) and accordingly has no difficulties with the peripheral variances described above. The latter model will be discussed in Section 1.4.2.

An associative model based on allophones

In his 1969 article, Wickelgren attacks a view concerning the processes underlying speech put forward by Lashley in his famous 1951 article. 'The problem of serial order in behavior'. In this article, Lashley shows that an associative model ('associative chain theory') for speech production is untenable. Such a model assumes that each element from a message (e.g. the phonemes of a word) triggers the following because of an existing direct association. The process may be guided by feedback from the on-going actions (Fairbanks, 1954), but it can also be located centrally which solves any difficulties that may arise as a result of long feedback loops. Lashley presents two arguments for the untenability of the model. In the first place, he points out that an associative model cannot solve the time-order problem, since the elements can and do occur in several different orders. From the example he uses in this context (the pronunciation of the word 'right' and its reversal 'tire') it can be inferred that the elements are conceived of as phonemes. The second argument refers to retroactive coarticulation. Though an associative model can explain effects of preceeding events, it is hard to see how it can anticipate future events. Besides, as Lashley remarks, serial actions are very easily executed in different rates. From these facts, Lashley concludes that an 'associative chain

theory' is not adequate, instead, he proposes a relatively independent mechanism that determines the serial activation of motor elements.

Wickelgren argues that the question whether an associative model is tenable or not depends on what is considered to be the behavioral unit. Instead of Lashley's (context-free) phoneme, he proposes a (context-dependent) allophone as unit. Thus the word 'stop', for instance, is not composed of the elements s, t, o, p but of the elements $_{\#}S_{t}$, $_{s}T_{o}$, $_{t}O_{p}$, $_{o}P_{\#}$. When a person wants to pronounce the word 'stop', it is assumed that the auditory representation $_{\#}S_{t}$, $_{s}T_{o}$, $_{t}O_{p}$, $_{o}P_{\#}$ calls up the corresponding elements in the articulatory domain, which are not as yet in any sequential order. The correct order is brought about by associations between the elements, whereby the association of $_{\#}S_{t}$ is greatest with $_{s}T_{o}$, and the association of $_{s}T_{o}$ is greatest with $_{t}O_{p}$, etc. Therefore $_{\#}S_{t}$ evokes $_{s}T_{o}$ and $_{s}T_{o}$ evokes $_{t}O_{p}$ and so forth. The fact that the word 'stop' can also be pronounced in a different order, for instance 'post' does not pose a problem for this model, since the two words have different internal representations. According to Wickelgren, the model cannot deal with words in which two identical pairs of 'elementary motor responses' are followed by a different 'elementary motor response'. As an example he gives the word 'lampblack' = /lampblak/ and remarks: 'This sequence has two 'identical' pairs of adjacent phonemes followed by a different phoneme in the two cases /la/ followed by /m/ and later /la/ followed by /k/.' (p. 6). Here the model would fail. However, says Wickelgren, there is no need to reject the theory, for in the first place only very few words have the mentioned properties, and secondly, the problem can be solved by assuming that these words are cut in two and each part treated separately. A better solution according to Wickelgren would be to suppose that besides context-allophones, there are also stress-allophones 'Thus the two /a/s in /lampblak/ are not identical. they differ in stress' (p. 7 op. cit.).

These ad hoc modifications, however, are not necessary since the two elements already differ in the unmodified conception: they are namely coded as $_{l}A_{m}$ and $_{l}A_{k}$. Consequently, the model works well in these cases.

There are cases where the model is ambiguous. Take for instance the artificial word 'OTOTOT'. The internal representation will be $_{\#}O_{t}$, $_{o}T_{o}$, $_{t}O_{t}$, $_{o}T_{o}$, $_{t}O_{t}$, $_{o}T_{\#}$. In this word, the associative strength between the first T and the second O will be as large as the associative strength between the first T and the third O. Therefore, one might expect the res-

ponse OTOTOT as well as the response OTOT. In this context we mention a study by McKay (1970) showing no higher frequency of occurrence of the speech errors predicted by Wickelgren's model.

The following objections can be made against the model. In the first place, the number of elements that is assumed (and accordingly the number of internal representations) is near 10^6*. Wickelgren brushes this point aside by remarking that the number of neurons in the brain amounts to approximately 10^{10}. Though this may be true, one should check whether or not a more economical solution is possible. Secondly, there is no evidence to support the existence of allophones as independent units in production, whereas such evidence has been found in support of phonemes. A final objection comes from McNeilage (1970) who interprets Wickelgren's model as a speech production model in which the articulatory elements are conceived of as motor commands. Such a 'motor command model' in which the possible responses consist of a limited number of motor commands cannot explain how a person is able to speak with a distorted speech apparatus (for instance with the teeth clenched).

* *This number is based on computations which assume that, besides stress-allophones, there are context-allophones caused by influences from adjacent phonemes. The above-cited study by Sussman et al. (1973) demonstrated, however, that there are also allophones caused by effects from more distant segments. This implies that Wickelgren's estimate is too low.*

1.4.2. An alternative model: invariant target positions

A model that assumes that each phoneme is always realized by means of one target position for the articulators, is more economical than the flute-playing model. The latter model indeed supposes that each sound can be realized by more than one finger position (= target position). Another point of difference may appear to exist in the manner in which the target positions are actually realized.

The idea of one target position per phoneme is, in part, based on investigations by Lindblom (1963) who made a spectrographical analysis of vowel reduction. Vowel reduction is a tendency to pronounce unstressed vowels as schwa. From Lindblom's study it appears that reduction is most probably determined completely by reduction of duration (increase of stress is accompanied by lengthening of segment duration). See Figure 7.

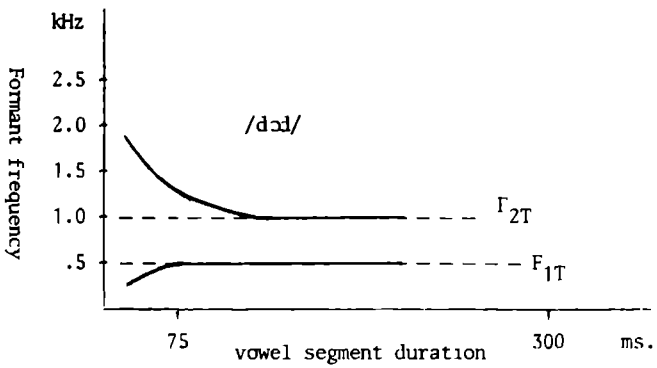

Figure 7. Formant position as function of segment duration. F_{1T} and F_{2T} are target positions of F_1 and F_2, respectively. (After Lindblom 1963)

From the spectral analysis of several CVCs differing in stress (i.e. in duration), it appears that the degree to which the formants of a spoken vowel actually reach those of the 'ideal target position'

(i.e. the positions of the formants in the steady-state part of a sustained vowel) can be predicted by means of a function that relates formants to segment duration. The asymptotes of the separate (exponential) functions for F_1 and F_2 form the ideal target positions of both formants. The functions can be specified independent of phonetic context and can thus form an invariant characteristic of a vowel. This finding fits a speech-production model that issues 'go-to' commands corresponding to phonemes to the articulators. Such a model assumes that peripheral coarticulation may be the result of inertness in the peripheral apparatus, such that when a command for the next position arrives, it has not yet attained the ideal target position of the previous command.

Validation of the hypothesis that there is only one 'go-to' command per phoneme cannot be based on electromyographic data, but must rely on an analysis of the movements of the articulators themselves (direct measurements, cinefluorograms, etc.). This testing is hindered, however, by the assumption that the target position specified in the 'go-to' command need not actually have been reached, due to the inertness of the peripheral system. For this reason, the reduction phenomena cited above are no counter-argument against the hypothesis. Below, we shall discuss some phenomena that appear to be incompatible with the assumption of phoneme invariant target positions.

McNeilage and De Clerk (1969) found in cinefluorograms they made during the pronunciation of CVCs of the type /gVg/ that hyoid elevation during the production of the final /g/ was smaller when the preceeding vowel was low /ɔ, æ/, than when this was a high vowel /u,ɪ/. The speech sound under consideration, the final /g/, is the last of a sequence. This means that there is ample time for the realization of it and accordingly there is no reason to suppose that the same articulatory position should not be reached in all cases. Thus it seems that different commands are used for the production of /g/ in different contexts.

Öhman (1966) reported a similar phenomenon. He showed a coarticulation effect in X-ray motion pictures, namely that the dorsopalatal area is smaller during the production of /u/ before /g/, than during the production of /u/ before /d/ (See Figure 8, page 26). It appears that this phenomenon remains even if the /u/ is lengthened considerably. This finding should point to an interpretation, not in terms of an overlap of a 'go-to-/u/' command, by a 'go-to-/g/' or by a 'go-to-/d/' command, but in terms of two different commands for two distinct articulatory positions, namely /u/ before /g/ and /u/ before /d/.

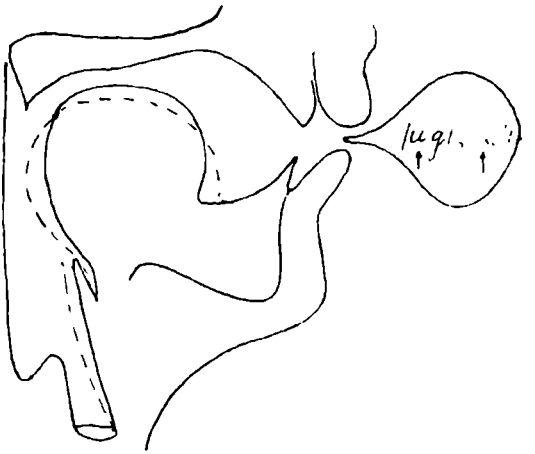

Figure 8. Tongue position during the production of /u/ before /g/ (solid line) and /u/ before /d/ (dashed line). From Öhman (1966).

McClean (1973) observed a considerably delayed onset of velar movement in those cases where a marked junctural boundary existed between the two vowels in a CVVN (N = nasal) sequence. Also this observation is incompatible with the assumption of a phoneme invariant 'go-to' command.

Sussman et al. (1973) described a phenomenon that was termed 'neuromuscular compensation' during the production of V_1CV_2s. If the C is a bilabial, this sound may be coarticulated by an open V_2, which consists in a decreased elevation of the lower jaw. This reduced jaw movement is compensated by the underlip which is raised more than normal (this enhanced activity can be found in the electromyogram of the M. mentalis) in order to produce the necessary bilabial constriction. The demonstrated neural compensation will not be expected from a phoneme-invariant 'go-to' command model.

Abbs (1973) examined the consequences on speaking of anaesthetization of the N. mandibularis. In this investigation an effect was found that resembles that of the preceeding study. It appears that in the nerve-block condition the jaw-closure movement for the production of the final /p/ in /p æ p/ is too slow or insufficient. The author concludes that other articulators (most likely the lips) have compensated for the insufficient jaw movement. This study differs from the above mentioned research

by Sussman and others in that it describes the speaker's apparent ability to compensate for occurring distortions in the articulatory system. In this respect it is related to those studies that examine the consequences of artificial distortion or constraints of the vocal tract on articulatory behaviour. This research in particular yields data incompatible with the 'go-to'-commands model.

Lindblom and Sundberg (1971) described an experiment in which subjects spoke a number of sustained vowels while clenching a small block between their teeth, thus fixating the jaw opening. A spectrographical analysis was made of the first glottal pulse, thus eliminating the possibility of the use of acoustic feedback (the only type of feedback from which an adequate correction might be inferred) for readjusting the articulators. From this analysis it appeared that the formant patterns approach the normal values within rather narrow limits. These spontaneous articulatory compensations can only be interpreted by assuming that the speaker can make use of alternative target positions.

This experiment is in agreement with the everyday observation that people can speak with a pipe (for an investigation on 'pipe' speech see Nooteboom and Slis, 1970), pencil or potato in their mouths or with self-imposed constraints, such as clenching the teeth or not rounding the lips. Also in these cases very complicated reorganizations of the target positions of the articulators have to be realized. In these cases some form of practice (making use of feedback) may be involved.

The phenomena cited above cannot be explained by a model of speech production that assumes phoneme invariant 'go-to' commands. This at least implies that the model in this form is too simple. The most serious objection against such a model is that it is too rigid. The cited phenomena do not at all indicate a rigid operation of the articulatory system in terms of invariant commands, but on the contrary, an extremely functional and adaptive mechanism which exploits the possibilities of the vocal tract by taking into account perceptual requirements on the one hand and the mechanical constraints of the oral structures on the other. Especially the ability to compensate for distortion – behaviour that may be compared to the transfer of writing with the one hand to writing with the other – point at the great pliability of the system. It also indicates that the assumption of motor-fixated neural commands corresponding with phonemes is untenable. If we still wish to use the term 'target' it will have to be defined more abstractly. We cite McNeilage (1970): 'It is difficult to believe that a speech production system

based on storage of discrete movement patterns could make such a spontaneous adjustment (needed in compensating oral distortions - author's note) by immediately producing a new set of motor patterns.' (p. 189). McNeilage then tentatively supposes that for each phoneme there are a limited number of articulatory targets, specified in an internalized spatial coordinate system, which represents the vocal tract. The latter idea stems from Lashley (1951). If it is true, there is still the question of how the speech-production system decides for a certain target position.

In the flute-playing model we have not as yet paid attention to this aspect. It will supposedly be clear that in flute playing this will be less of a problem than in speaking. First, the number of alternative finger positions is limited and explicitly specified. Thus it does not appear too difficult to describe the choice of a target position with the help of a limited number of rules that are based on some insight into the operation of the flute on the one hand and on the perceptual acceptability of the corresponding sounds on the other. Secondly, there will usually be no need for the flute player to adjust to occurring distortion in his flute, except for very simple distortion in one dimension. Most likely the flute player will not be able to compensate for more complicated structural distortions of the above mentioned type that can be applied to the vocal tract. To start with the latter point: the observed fact that a speaker knows how to produce more or less adequate speech sounds even with a severely distorted articulatory apparatus points towards the very special way in which the articulatory apparatus must be internally represented. This 'knowledge', at any rate, includes more than a limited number of target positions required for the production of a set of sounds.

The choice of a particular way of realizing a speech sound in a certain context must be based on the knowledge as to which 'allophone' will be perceptually acceptable within the given context. This 'allophone' could be conceived of as a point in an multidimensional perceptual space that belongs to a cluster of points forming all possible 'allophones' of one phoneme. The corresponding articulatory position might then be found by mapping the perceptual space on the internal space representing the articulatory configurations. A very similar point of view has been brought forward by Nooteboom (1970).

The presented description is besides vague also not quite adequate: in the case of a distorted vocal tract, the choice of an 'allophone' in

the perceptual space cannot be made independent of the existing distortion, which implies that somehow the altered properties and possibilities of the vocal tract are 'known' in the perceptual space. It seems impossible at present to describe this mapping process in a model that is sufficiently concrete.

1.5. The role of feedback

In the preceeding text it could be shown that the flute model gives an adequate description of some important aspects of speech production. The model, however, still needs some elaboration. Besides the aspect discussed in the last paragraphs - the determination of allophones and its articulatory counterpart – there remains the stage in the process in which the necessary actions for the realization of an articulatory position are determined and the possible role of feedback in this part of the process. This latter aspect will be discussed below in some detail.

Recently, a great interest has been shown in the possible role of feedback in speaking, or, in other words, for the question as to whether speaking must be considered an open-loop rather than a closed-loop process. This interest is apparent from investigations in which feedback is actually experimentally manipulated as well as from studies that, being based on the former research, try to develop a view on the role of feedback. Before discussing this research we shall first take a look at the closed-loop model of speech production by Fairbanks (1954) which may be considered representative for the type of models that were based on the science of cybernetics, then in development.

Fairbanks's model

In his 1954 article, 'A theory of the speech mechanism as a servo system', Fairbanks proposes a theory of speech production in which he hypothizes that afferent information on the process of speaking that reaches the subject via acoustic, tactile and kinesthetic pathways not only has a monitoring function, but also plays an important role as feedback in a closed-loop process for the production of speech. Via the mentioned channels, the feedback is fed to a central comparator where it is compared with a norm. This comparison results in a correction signal (error signal). If the correction signal is zero (i.e. input matches output), the next 'speech-unit' is selected and sent to the effectors (see Figure 9, page 31).

Fairbanks can only propose this model as a speech-production model because he conceives speaking as a tracking task (he actually makes a com-

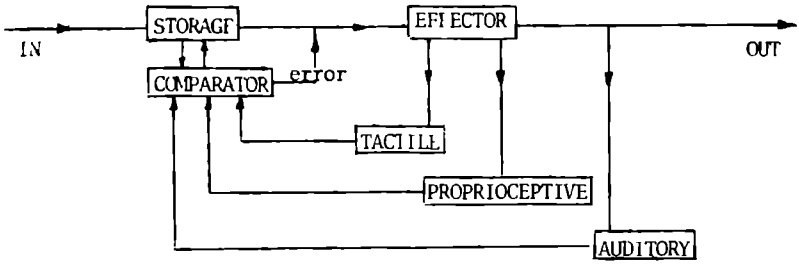

Figure 9. Simplified model of speech production by Fairbanks.

parison with a thermostat) during which correction can be carried out on the output until some defined norm has been reached. Such a model, however, does not sufficiently take into account the complicated time order requirements in the speaking process, which are the very narrowly defined temporal structure of speech on the one hand and the duration of the feedback loops on the other. Because of these properties, it will in many instances not be possible to carry out correction responses on a speech unit, for the simple reason that at the moment feedback arrives centrally the norm will already have been replaced by the norm belonging to the following speech unit. Let us have a closer look at the auditory feedback, the type of feedback that according to Fairbanks plays such an important role in the control of articulatory behaviour.

The reaction time (RT) to a simple auditory stimulus is approximately 130 ms., but when acoustic information is used for a correction response, reaction time is estimated to be much longer. Simon (1968) examined reaction time in a two-choice RT-task where the subject had to move a control handle to the left or to the right in response to corresponding verbal commands. In the most favourable conditions the investigator measured an average reaction time, defined as the time elapsed between a verbal command and the beginning of a movement, of 370 ms.! We may reasonably assume that at least the same amount of time is required for the formation of an articulatory correction response, because complicated transformations must be executed in order to translate a discovered deviation in the acoustic signal into an adequate correction command to the articulators. There will certainly be 'speech units' that are executed within this period of time, so that correction based on auditory feedback is ruled out. This also holds, even when it

is assumed that the system is able to predict what state will be attained in the near future, since these predictions can only be based on actions corresponding to the realization of the ongoing speech unit.

Fairbanks's conception of a completed speech unit that triggers the following speech unit is an idea that fits an associative model. The conception, however, only makes sense if the speech unit is defined. If a speech unit is to be considered an acoustic segment, it should be taken into account that, as Lenneberg (1967) has shown, acoustic units (segments with a defined duration) do not remain units (segments in the time domain) after transformation into neural commands, partly because of differences in length of nerves to the various muscles essential to articulation (activation latency). In fact, a considerable rearrangement and overlap of neural commands belonging to different acoustic segments must be supposed to occur. Fairbanks remains very vague about the content of the speech unit, he writes 'such a control unit should not be identified with any of the conventional units such as the phoneme, the syllable, the word or the word group. .... It might be ventured tentatively that the unit of control is a semiperiodic, relatively long, articulatory cycle, with a correlated cycle of output.'

However much the model is still in want of further elaboration, it does in any case predict disastrous effects on speech production if feedback is eliminated or reduced. This topic will be examined below. Mysak (1966) and Chase (1967) developed models for speech production that are essentially identical to Fairbanks's model.

Empirical evidence of feedback deprivation

A speaker receives feedback about his speaking behaviour via three sense modalities auditory feedback; tactile feedback concerning contacts between the different articulators; and kinesthetic feedback about stretch and movement of muscles and joints. Vibrational feedback will be left out of consideration because of its unimportance.

Research into the effects of experimental manipulation of feedback on speaking, can accordingly be devided into studies on the effects of delayed auditory feedback (DAF) and the effects of removing or severely limiting acoustic feedback (masking noise); studies on tactile feedback deprivation by anaesthetizing oral surfaces; and studies on combined tactile and kinesthetic feedback deprivation by blocking afferent nerves from one or more articulators. First we will report in brief a number of these

studies and subsequently evaluate them.

In the well-known studies of the DAF effect on speaking, the subject receives his own acoustic speech signal delayed via a headphone. The first investigations were done by Lee (1950 a, b, 1951) followed by a large number of studies by other investigators. Review papers have been written by Smith (1962) and Lansener & Lansener (1963). The most important effects reported by all authors are slowing down of speaking rate, increase of loudness and pitch and growth in number of speech errors. Lansener & Lansener (1963) mention that their subjects speak in a more monotonous fashion under the experimental conditions. Chase (1967) and Lansener & Lansener (1963) report very considerable interindividual differences in subjects, some subjects do not even show any effect at all of the feedback delay. The effect is greatest when the delay lies between 200 and 250 msec. (Fairbanks), delays smaller than 100 msec. and larger than 1 sec. do not cause any trouble.

Fairbanks (1955) and Lansener & Lansener (1963) made a detailed analysis of speech errors under DAF conditions. According to the latter authors the most frequent errors are, in order of importance repetition of syllables, substitutions of words and syllables, omissions of words, syllables and speech sounds. Fairbanks (1955) supposes that DAF conditions mainly affect the formation of articulatory movements.

The effects that occur in DAF conditions in fact tell us little about the consequences of elimination or reduction of feedback, but more about the consequences of supplying subjects with incorrect feedback. This fact makes the interpretation of the observed facts very difficult, actually the proposed interpretations of both Lee (1950) and Fairbanks (1955) are inadequate (ref Smith 1962). On the question concerning the function of auditory feedback in speaking, the most important finding is that all subjects continue to speak (be it less well) under experimental conditions and that some subjects appear to be quite resistent to feedback delay. Clearly, these data do not allow the conclusion that auditory feedback plays an indispensable role in speech production.

In a more direct attempt to establish the function of auditory feedback in the process of speaking, Ladefoged (1967) presented to both ears of subjects white noise that was so loud that the subjects could no longer hear their own voices, not even by bone conduction. The subjects made spontaneous remarks. The following pecularities were found in their speech. many vowels were changed with respect to length and quality, some non-nasal speech sounds were nasalized and vice versa. The latter effect was also reported by Rutherford (1967) who related the finding

to the small sensitivity of the velum and the accordingly poor tactile feedback from this articulator. Other effects are concerned with pitch and intonation: the pitch was found to be higher, the intonation more monotonous with a preference for a simple falling intonation pattern. Ladefoged reports that the subjects could sing a familiar tune. Rutherford, however, asserts that subjects can hardly sing under the circumstances. Because of tightening of the muscles of the larynx, changes in voice quality occurred. Ladefoged notes that all effects mentioned are not necessarily caused by the elimination of auditory feedback, but can perhaps be attributed to the effect of the very disruptive strong noise on speaking. Ringel and Steer (1963) also report that their subjects spoke louder and slower; moreover they observed changes in the phonation-silence ratio. Maybe these latter effects also are secondary. Much more research is required to solve all problems connected with the function of auditory feedback in speaking.

More research has been done on the consequences of deprivation of tactile feedback of the vocal tract on the speaking process. Before describing the results of this research we shall present some notes on the manner in which the tactile feedback is eliminated. Normally, one of two methods is followed: either the surfaces of the oral cavity are anaesthetized, or one of the sensory cranial nerves serving the mouth is blocked. A problem inherent to the latter method is that besides the tactile sense also the kinesthetic sense may be blocked, which makes a differential analysis of the contribution of the two senses impossible. Ringel and Steer (1963), Scott and Ringel (1971) and Horii et al. (1973) anaesthetize the N. mandibularis in order to deprive the subject of tactile feedback. Scott and Ringel do report that it cannot be stated with certainty whether interference thus also occurs in proprioceptive (kinesthetic) afference. According to the latter authors, the afference from the muscle spindles of the tongue is ascribed to different nerves, the mandibular nerve being one of them. Abbs (1973) also anaesthetized the mandibular nerve but his purpose was not to examine tactile feedback deprivation, but to check the consequences of blocking the gamma system, which is a part of the kinesthetic system. It should be noted that Abbs used a special procedure which enabled him to investigate only the sensory aspect of the nerve and not the motor (the mandibular nerve is a subdivision of the trigeminal nerve that also contains motor fibres to the masticatory muscles). He only starts to measure the effects of the nerve block if and when the motor innervation appears to be restored (the alfa motor neurons recover sooner

from the anaesthetization than the gamma system). It is surprising that with respect to this problem only one reference has been made in the three abovementioned papers. Scott and Ringel state that the motor innervation of some muscles involved in speaking may also be blocked. Still they suppose, however, that the effect on speaking will be negligible.

It follows from the preceding discussion that one should be careful in interpreting the effects of nerve-block anaesthetization. not only tactile feedback, but also proprioception and possibly motor innervation may be affected. We will now turn to the results.

Ladefoged (1967) reports that subjects who had the surface of the inside of the oral cavity anaesthetized, had difficulty in articulating sounds involving lipmovements /p,b,n,f,v/ or tongue movements /s,z,d, t,l/. Ringel and Steer (1963) also mention articulation errors when local anaesthetization of the oral cavity was applied, a considerable increase in errors was observed when the anaesthetization was realized by means of nerve block. Articulation errors were determined by judges that counted the number of errors they heard. Horii et al. (1973) report a frequency decline of approximately 3000 Hz (!) in /s/ sounds produced by subjects deprived of tactile feedback.

In a recent study, Scott and Ringel (1971) made a detailed analysis of articulation errors after elimination of orosensory perception by means of anaesthetization of several oral nerves. They could observe that the majority of articulation errors are made up of non-phonemic errors such as: loss of retroflexion and lip-rounding, decreased fricative constriction and retracted place of articulation.

For the rest it needs to be mentioned, that speech produced under oral anaesthetization remains very intelligible (Horii et al. 1973), and that a combined elimination of both auditory and tactile (and kinesthetic) feedback does not result in a dramatic deterioration as compared with the condition in which only tactile feedback has been ruled out (Ringel and Steer 1963).

Several authors (Liberman 1967, McNeilage 1970, Sussman 1972) have assigned an important role to the gamma system for the realization of 'go-to' commands without indicating, however, precisely how this function is achieved. The gamma system is a second muscle-innervation mechanism, that does not act on the motor end plates as does the alpha system, but on muscle spindles within the muscles. Afferent fibres from the nuclear of the muscle spindle form a reflex arc with alpha motorneurons. If the muscle spindle is stretched either because of decontraction of the muscle

body, or because of contraction of the muscle spindle itself via gamma efference, the sensory nerve endings fire. Via the reflex arc with the α motorneuron this results in muscle contraction until the stretch in the muscle spindles has been eliminated. This description of the gamma system is considerably simplified, for more detailed discussion see for instance Bell (1970), Abbs (1973).

The description given suggests two possible functions of the gamma system. One function could be to compensate for external forces that change muscle length. A second function could be the 'automatic' realization of a predetermined muscle length.

The role of the gamma system has actually been demonstrated in maintenance of posture. It also appears to be involved in the development of speed in the initial phase of a movement (via activiation of the alfa motorneuron) and in the slowing down of speed in the final phase to prevent overshoot (by means of antagonistic facilitation). The latter function is also demonstrated in voluntary movements as for instance in throwing darts (Mortimer and Aakert 1961). Abbs (1973) in his abovementioned study, found that blocking the gamma system of the jaws results in a consistently smaller jaw opening during the production of open vowels as compared to a control condition. This effect may again be attributed to an insufficient development of speed in the initial phase of the movement.

The second function suggested above, namely the possibility to realize 'go-to' commands automatically, might be worked out as follows the target position is put via the gamma-efferent system into the muscle spindle which sees to it via the reflex arc that the appropriate target position is realized. According to this view the actions to be performed for the realization of a particular target position are not computed centrally, either with or without information from the periphery – as is assumed in the flute-playing model – but the target position is realized automatically. It needs mentioning that such a function of the gamma system has not been suggested in the physiological literature except in a not very explicit manner by Pribram (1971).

Finally, central afference has also been demonstrated in the muscle spindle (Abbs 1973), so that a closed-loop system based on kinesthetic information can be conceived of analogous to the auditory and tactile feedback loop. Because of the very high transmission time in muscle spindle afferents from the tongue (4-5 msec.) Sussman (1972) argues for a closed-loop model of speech production. It should be noted, however,

that this duration applies only to the afference time and not to the total loop time which is estimated to be considerably longer: 20-80 msec. Abbs (1973) comments on this time constant: 'Since the interval required to attain a significant change in tension will probably double this time constant, it appears unlikely that feedback correction can be employed once such rapid movements are initiated.' (p. 180).

There is as yet not enough empirical evidence to enable us to determine the existence of the hypothesized functions of the gamma system. From the abovementioned work by Ringel and Steer (1963), Scott and Ringel (1971) and Abbs (1973) who all report that blocking the gamma system does not lead to an inability to speak (or to bring the jaws in the necessary positions for speaking) it must be concluded anyway that the view of the gamma system as an automatic position-realizer is too simple.

Moreover it follows from these investigations that central afference from the muscle spindles is not absolutely indispensable in the process of speaking. The above discussion confirms us in our view that the reported findings are quite incompatible with a completely closed-loop model of speech production as the model by Fairbanks supposes. On the other hand, a completely open-loop model predicts that no errors will occur under feedback deprivation conditions, which is clearly not the case. Neither do the reported errors (because of their modality specificity) fit a model in which the feedback only has a monitoring function.

We tentatively conclude that a true model of speech production cannot be completely open-loop although it appears that the subject can be deprived of a great deal of feedback even from several sense modalities at the same time without losing his capacity to speak intelligibly.

A more exact determination of how and when sensory feedback functions within the process of speaking must be provided by future research.

Resumé

In the preceeding chapter, a model has been proposed for the description of a number of aspects relevant to the process of speaking. In this model, the process of speaking is compared with the process of flute playing of which a tentative description is given.

In a later section, two alternative models of speech production are

discussed which both assume that some one-to-one relation exists between phonemes and neural commands for the production of these phonemes. The first model supposes phoneme invariant motor commands to the articulation muscles, the second model does not state that invariance exists up to the level of muscle innervation, but it assumes that each phoneme has one corresponding 'go-to' command in which a target position for the articulators is specified. The models are confronted with conflicting data from electromyographic and cinefluorographic investigations into the process of speaking. Thus it is concluded that both models are untenable, and that phoneme invariance cannot be reduced to motor invariance. Apparently, phoneme invariance only exists in the perceptual domain.

The furnished data seem to support the proposed flute-playing model of speaking in which speech is conceived of as a sequence of sounds produced by an instrument characterized by continuous transitions from one state to another. Specifically, the flute-playing model firstly assumes that several different articulatory configurations are possible for the production of the same phoneme; for each actual production a choice has to be made of one specific oral configuration which will, among other things, be based on the phonetic context in which the sound appears. Secondly, the model supposes that quite different actions are required for the production of a specific oral configuration, depending on the articulatory position of the preceeding phoneme. Thus it follows that one can only consider a speech sound as being mastered if the subject can produce it in all occurring phonetic contexts.

In a final section, the role of feedback in the process of speaking is examined along with the question whether speaking should be considered an open-loop or a closed-loop process.

The presented view on the process of speech production has several interesting implications for different areas of speech learning.

1. The model suggests an explanation of the as yet ill-understood discrepancy between babbling and verbal utterances. It has been reported in the literature that a child can sometimes produce a particular sound in a babbling utterance, whereas at the same time he appears to be unable to produce the sound in a verbal utterance. This discrepancy might be attributed to the fact that the child can produce the sound only in specific articulatory sequences but not in others.

2. Another implication of the model lies in the field of the assessment of articulatory performance. Articulatory performance has often been

assessed by determining the frequency of occurrence of a specific speech sound. Some authors (e.g. Snow, 1963, Olmsted, 1971) have in their analyses also taken into account the position of the speech sound in the utterance (without explaining, however, why this was done). Our model suggests that phonetic context is a very important variable that has to be taken into account when interpreting acquisition data It is therefore proposed that acquisition data are reanalysed now as a function of their phonetic context, especially the preceeding speech sound Then it might appear that a particular sound appears consistently earlier in some context than in other. From a preliminary examination of speech errors of deaf boys it appears that an analysis as suggested above makes sense.

3. Finally our model suggests that a phoneme must be practised in the various contexts in which it may appear. Moreover, from an analysis such as mentioned above, an order, or perhaps a hierarchy of difficulty of pronouncing phonemes in different sequences might be inferred. That would be of great value not only to the practice of speech training, but also to its theoretical understanding.

2. SPEECH CORRECTION

Our first chapter dealt with processes underlying the articulatory aspect in speech production. A model was developed in which the most important processes were described. The present chapter will look into speech correction. Two methods of correcting speech will be compared, their advantages and disadvantages weighed, and the implications for speech training in general and speech correction for the deaf in particular will be indicated.

Before continuing, it will be advisable to comment on the distinction between speech acquisition and speech correction. The acquisition of speech is an integral part of the development of a child learning to speak. It cannot be separated from the development of language, for the child that learns to speak is also learning a language. This implies that he is assimilating a complicated system with many subsystems, such as semantics, syntax, morphology and phonology. Articulation is only one subsystem of the complex whole. Pronunciation errors made in the acquisition period may be jointly due to problems of acquisition at any of the distinct levels. This greatly hampers an understanding of the development of the different subsystems, in this case, articulation. In this context, Winitz (1969) makes a relevant distinction between phonemic and phonetic acquisition 'Phonemic acquisition involves the learning of the phoneme system of the community language – the functional units of the language that signals semantic distinctiveness. It involves the learning of the acceptable phoneme sequences of the language...' (p. 65). By phonetic acquisition he means the development of the capacity to make the required motor responses. Winitz tries to make it plausible that many pronunciation errors children make during acquisition can be ascribed to an insufficiently developed phonemic system. Though this may be so, especially in the early stages of development, we still believe in a certain independence between acquisition of language and acquisition of articulation. This independence is demonstrated, not only in the phenomenon that a passive command of language generally preceedes an active command (Jakobson 1968), but also by patients suffering from speech difficulties. Lenneberg (1962) described a patient with congenital anarthry (inability to produce any sound) who nevertheless had a good understanding of language. A similar

case described by Chase (1967) concerns a female patient with serious congenital speech difficulties but otherwise normal speech perception.

We shall not now go into the characteristics of speech development in the prelingual and lingual stages. This process of development has been treated in greater or lesser detail by many authors and, depending on their personal field of interest, placed by them in the framework of psychological theories of learning, or in the framework of linguistic theories, viz. transformational generative grammar (e.g. Gregoire 1937, Jakobson 1968, Skinner 1957, Caroll 1960, McCarthy 1966, Smith and Miller 1966, Lenneberg 1967, Berry 1969, Winitz 1969, McNeill 1970, Braine 1971, Menyuk 1971, Olmsted 1971, Kremers 1972, Schaerlakens 1973, Smith 1973).

In practice, a need will be felt for articulation correction if, within the whole development of speech, there is a lag in one subsystem, in our instance, the motor development required for faultless speech. In the deaf too, where development of the different subsystems is presumably less well-integrated than in the hearing, there will often be cases of relatively isolated pronunciation problems Solution of these problems is the object of the correction method that we shall discuss below. It will appear, finally, that the distinction between acquisition and correction is by no means a sharp one The most important distinction lies in the separate treatment of one specific facet.

We feel that the possibility of regarding articulation as a relatively independent aspect of speech acquisition justifies us in using the skill model. That speaking is analogous to a motor skill and learning to speak analogous to the acquisition of a skill is a view we should like to make a comment on. Lenneberg (1967 p 132 et seq.) has pointed to a number of differences that are supposed to exist between the acquisition of speech and the acquisition of a sensory-motor skill. Speech does not display the considerable interindividual differences that are met within the motor skills, the beginnings of speech always occur at approximately the same age regardless of cultural differences or natural abilities, and it would seem that practice does not, or only slightly, affect the time of the onset of speech.

The above observations indeed emphasize the peculiar characteristics of speech development and indicate the importance in this process of maturation. However, these data should be cautiously evaluated. Granted, the majority of people have reached a level of speech performance which makes them intelligible for others sharing the same language,

2. SPEECH CORRECTION

Our first chapter dealt with processes underlying the articulatory aspect in speech production. A model was developed in which the most important processes were described. The present chapter will look into speech correction. Two methods of correcting speech will be compared, their advantages and disadvantages weighed, and the implications for speech training in general and speech correction for the deaf in particular will be indicated.

Before continuing, it will be advisable to comment on the distinction between speech acquisition and speech correction. The acquisition of speech is an integral part of the development of a child learning to speak. It cannot be separated from the development of language, for the child that learns to speak is also learning a language. This implies that he is assimilating a complicated system with many subsystems, such as semantics, syntax, morphology and phonology. Articulation is only one subsystem of the complex whole. Pronunciation errors made in the acquisition period may be jointly due to problems of acquisition at any of the distinct levels. This greatly hampers an understanding of the development of the different subsystems, in this case, articulation. In this context, Winitz (1969) makes a relevant distinction between phonemic and phonetic acquisition 'Phonemic acquisition involves the learning of the phoneme system of the community language – the functional units of the language that signals semantic distinctiveness. It involves the learning of the acceptable phoneme sequences of the language...' (p. 65). By phonetic acquisition he means the development of the capacity to make the required motor responses. Winitz tries to make it plausible that many pronunciation errors children make during acquisition can be ascribed to an insufficiently developed phonemic system. Though this may be so, especially in the early stages of development, we still believe in a certain independence between acquisition of language and acquisition of articulation. This independence is demonstrated, not only in the phenomenon that a passive command of language generally preceedes an active command (Jakobson 1968), but also by patients suffering from speech difficulties. Lenneberg (1962) described a patient with congenital anarthry (inability to produce any sound) who nevertheless had a good understanding of language. A similar

case described by Chase (1967) concerns a female patient with serious congenital speech difficulties but otherwise normal speech perception.

We shall not now go into the characteristics of speech development in the prelingual and lingual stages. This process of development has been treated in greater or lesser detail by many authors and, depending on their personal field of interest, placed by them in the framework of psychological theories of learning, or in the framework of linguistic theories, viz. transformational generative grammar (e.g. Gregoire 1937, Jakobson 1968, Skinner 1957, Caroll 1960, McCarthy 1966, Smith and Miller 1966, Lenneberg 1967, Berry 1969, Winitz 1969, McNeill 1970, Braine 1971, Menyuk 1971, Olmsted 1971, Kremers 1972, Schaerlakens 1973, Smith 1973).

In practice, a need will be felt for articulation correction if, within the whole development of speech, there is a lag in one subsystem, in our instance, the motor development required for faultless speech. In the deaf too, where development of the different subsystems is presumably less well-integrated than in the hearing, there will often be cases of relatively isolated pronunciation problems Solution of these problems is the object of the correction method that we shall discuss below. It will appear, finally, that the distinction between acquisition and correction is by no means a sharp one. The most important distinction lies in the separate treatment of one specific facet.

We feel that the possibility of regarding articulation as a relatively independent aspect of speech acquisition justifies us in using the skill model. That speaking is analogous to a motor skill and learning to speak analogous to the acquisition of a skill is a view we should like to make a comment on. Lenneberg (1967 p. 132 et seq.) has pointed to a number of differences that are supposed to exist between the acquisition of speech and the acquisition of a sensory-motor skill. Speech does not display the considerable interindividual differences that are met within the motor skills; the beginnings of speech always occur at approximately the same age regardless of cultural differences or natural abilities, and it would seem that practice does not, or only slightly, affect the time of the onset of speech.

The above observations indeed emphasize the peculiar characteristics of speech development and indicate the importance in this process of maturation. However, these data should be cautiously evaluated. Granted, the majority of people have reached a level of speech performance which makes them intelligible for others sharing the same language,

but if one observes accurately, one can establish enormous differences in the degree of perfection attained in speech. A large proportion of adult Dutch speakers fail to produce an acceptable /r/, others have difficulty in pronouncing the /s/. A number of recent pieces of research indicate a connection between an ability to recognize and compare objects placed in the mouth and an articulatory capacity such as the ability to form unknown sounds (Locke 1968, Ringel et al. 1970, Bishop et al. 1973). Typically, this is a connection one would expect to find in skills where sensory control always plays an important part both in acquisition and in performance.

Another phenomenon likewise typical for skill development, a division into stages (Fitts and Posner, 1967), has also been observed in articulation acquisition during the learning of a second language. Thus Kalikow and Klatt (1970) discovered that the subjects they selected for their investigation into pronunciation improvement only had pronunciation problems if they were required to produce certain sounds in running speech and not if they were permitted to pronounce the same speech sounds calmly in isolated words. Though these subjects gave the appearance of having attained a certain degree o acquisition, they had apparently not yet reached the stage of 'automation'. The above considerations give us grounds for also basing the discussion on speech correction on the skill model.

2.1. Speech correction methods

The adult speaker is able to produce the speech sounds of his language in the sequences that can occur in it. Sequences strange to his language will usually cause him some difficulty. This observation fits the model we have developed and described in Chapter 1.

From research done by Irwin (1941, 1946) the conclusion has been drawn that the babbling child's repertoire includes all possible sounds. This is only partly correct. In the first place, a child of this age produces all kinds of sounds with a certain unintention, analogous to the way a child can bang out notes on a piano, one cannot then consider there to be control over the production of these sounds. In the second place, as pointed out in the preceeding chapter, producing a sequence of speech sounds is different from chaining separate speech sounds there is no one-to-one correspondence between speech sounds and neural motor commands.

In producing words and sentences, i.e. sequences of speech sounds, a choice will have to be made out of a number of possible oral configurations; moreover, the speaker will have to determine how the chosen position will have to be reached from a given initial state. These are the problems that face the adult when he struggles to pronounce an unfamiliar sequence of speech sounds. These considerations should be borne in mind in the following discussion.

We shall discuss speech correction in the light of the flute-playing analogy and see whether there is agreement between the way in which a person learns to play the flute and the way in which he learns to speak.

Even a superficial consideration will show that there are two essentially different ways towards mastery of the flute with instruction and without instruction. It will appear that different processes are involved in the two manners. Both will appear to be relevant to a better understanding of speech acquisition and speech correction. The two ways, which we shall call 'learning through process shaping' and 'learning through output matching', respectively, will be dealt with in turn below.

Learning through process shaping

The most usual method of learning to play the flute is that where the pupil is taught which finger positions are used on his instrument. The finger positions can be given names or they can be indicated by means of notes. The pupil can now produce a sequence of sounds when the behaviour needed is specified in terms of finger positions to be attained at specific moments in time. Noteworthy is that it is up to the pupil to choose the actions which are needed in realizing the finger positions in the sequence.

A proportion of those who learn to play in this way are not able to form a conception of the acoustic correlate of the notation. A peculiar situation arises where the subject does not beforehand know what acoustic product will be generated. This he will hear only during and after the performance of a series of actions. There is for these persons thus no connection between the acoustic correlate of the product and the underlying behaviour needed to produce it. By following this method of acquisition, the subject has developed a very special 'knowledge' of the instrument, whereby the representation is coupled to motor behaviour and the sound is but a derivative. These people are therefore not able to play 'by ear'. Neither this situation, nor the way in which it has developed seem representative of the manner in which the hearing person learns to speak. In a number of respects it will appear to show similarities with the way in which the deaf learn to speak. We shall return to this presently.

Learning through output matching

The manner of learning a motor skill as described above is rather uncommon. It is restricted to those types of skills, such as flute playing, that are composed of a relatively small repertoire of specifiable actions.

In fact, learning to play the flute can develop in a quite different way, one which is probably the most natural. Here the pupil himself tries to discover how to produce tones and melodies with the aid of the instrument. Clearly this procedure shows more similarities with the manner in which the hearing learns to speak. Typically it may be said that the speaker who learns in this way can only 'speak by ear' only the auditory representation of a sound sequence enables him to evoke the necessary muscle movements for the pertaining sound sequence.

In contrast to the first method, there is here a direct gearing of perception and motor performance. Another characteristic feature of this method

is that the subject is hardly aware of the actions needed in the production of speech. It also appears, as was described in Chapter 1, that subjects who have learnt to speak in the latter way are able to cope with distortions of the speech organs, whereas for subjects who learnt via the first method this must be regarded as fundamentally impossible.

In order to avoid misunderstanding it must be stated that under normal circumstances, even if the first method is explicitly used in instruction, there is nothing that prevents the subject from enriching his acquisition by means of the second method, (i.e. by seeking and establishing connections between perception and motor performance), thus in practice, characteristics of the second method will be found even though apparently only the first method was followed.

We now wish briefly to explore the question of how this latter (natural) course of development can be described. It seems it could be typified by the subject's attempts to produce sounds and sound sequences that resemble sounds and sound sequences spoken around him. A number of authors (e.g. Allport 1924, Fry 1966) regard imitation indeed as the core of this learning process. Against this view objections have been made by authors who pointed out that a child does not really imitate (in a literal sense). Lenneberg (1964) pointed to the considerable acoustic (spectrographic) differences between the sounds produced by the adult and the imitations by the child. Beside this we would indicate the phenomenon that a child of this age omits, adds or changes speech sounds. These transformations seem to follow rules which are partly known (Jakobson 1968, Ingram 1974). Both observations argue that there is no literal imitation, but that between input and output complicated transformations occur. The most important objection against imitation as the explanatory principle of speech development is the fact that imitation presupposes the very mechanism whose development we wish to understand for, if a child is indeed able to imitate, it is able to resay heard speech sounds, in other words, it is already able to speak.

In the process of speech learning we can distinguish two components a perceptual and a motor. Perceptual development consists in the development of a capacity to discriminate relevant acoustic cues, which ultimately results in the recognition (identification) of the repertoire of sounds and the lawfulness (rules) governing the formation of sound sequences. Motor development may be defined as the development of the capacity to produce perceived elements and series of elements. The order followed here does not imply that in the actual growth of a child's speech, perception

first develops fully and only afterwards motor performance. On the contrary, the features of the development indicate a certain mingling . Yet we wish here to stress that perceptual development always runs ahead of motor development. Clearly, if the (perceptual) norm on which development is oriented is erroneously perceived, this will hamper development. (This view is diametrically opposed to the motor theory of speech perception.)

Older authors usually suppose that the adoption of a new speech sound in a child's repertoire will occur at the moment an agreement is discovered between a produced speech sound and a (possibly earlier) perceived speech sound in his environment. In this view, the production of speech sounds is regarded as accidental. Allport (1924) also supposes that the child only has at his disposal those sounds that he once chanced to make.

Psychological learning theory considers articulatory development as being determined in particular by mechanisms described in the framework of conditioning theories, such as reinforcement and shaping. We shall briefly discuss this view because it might provide perspectives for speech training, seeing the mechanisms mentioned make possible the active manipulation of behaviour.

Particularly investigators such as Skinner (1957) and Staats and Staats (1962, 1963) see the development of articulation as a conditioning process. correct speech attempts are reinforced by the parents and so are approximations of correct speech sounds (at least sounds that are regarded by the parents as approximations) so that a kind of shaping may be brought about.

Here, we will mention some objections against this view on the learning process. Firstly there are a number of observations that do not seem to fit into a conditioning model.

1. Parents often reward definitely incorrect sound productions, especially during the initial stage of the development

2. Articulatory development does not appear to come to a halt when the child is for some reason or other not able to utter and practice speech sounds during a certain period. Lenneberg (1967) described a case of a patient that was tracheotomized from 8 to 14 months of age. During this period, the child cannot, of course, be reinforced and shaping of attempts cannot take place. Yet it appears that the child, from the moment he is again able to vocalize, generates the speech sounds that are typical for his developmental stage.

3. The same applies to children who have been seriously neglected or

have been hospitalized during a certain period. In such a period, these children hardly ever speak, however, as soon as they are given due attention, it appears, that they did learn during this period in which they did not vocalize. It is said, in this context, that the active command of language lags behind the passive. This is, in fact, not true. As soon as the child is placed in an environment, in which he feels that speech or speech attempts are meaningful, he appears to be able to perform these.

4. A similar phenomenon is observed in the fact that there are substantial differences in the degree and the duration of babbling by children in their first years of age, whereas these differences are not reflected, as it seems, in their speech development

A theoretical objection against the conditioning model is that it only gives a formal description of the consequences of reinforcement and that it does not say anything about the processes that underlie the sound generation itself. Silently or sometimes explicitly it is assumed that a child must at least once accidentily have produced a speech sound in order for it to become included in the speech repertoire. This trial and error principle is contradicted by the fact that the development of the speech-sound repertoire in the prelingual as well as in the lingual stage appears to be highly structured (Lewis 1963, Fry 1966, Wellman 1931, Poole 1934, Templin 1966).

From the foregoing, we conclude that speech acquisition cannot be adequately described as a process guided by conditioning. We do not want to assert, however, that conditioning does not play any role in speech acquisition (especially for the motivational aspects it is of importance) but we want to stress that it does not satisfactorily describe the essence of the developmental process under consideration. It is therefore a pity that many textbooks (also recent ones e.g. Winitz 1969, Eisenson 1963, Berry 1969) give almost exclusive attention to conditioning techniques as an aid to speech training. The most important task of the speech trainer is to elicit desired articulatory responses from the pupil and to consolidate them. The methods to arrive at these aims must be based on insight in the processes that underlie normal speech production and its development.

Two earlier mentioned observations throw a special light on the process of speech learning. The first is the observation that practising seems only of very little importance for speech acquisition. The second is related to the observation that the adult can produce intelligible

speech with a severely distorted vocal tract (an activity that can be produced by everybody without apparent practising).

From these observations one might reasonably assume that the essence of the process of speech acquisition consists in a perceptual development. Consider again the second observation. From it might be inferred that, if the correct perceptual representation is available, the formation of an adequate speech sound (which means in this case. optimal within the possibilities given) does not appear to give much trouble, even when the new configuration differs considerably from the ones normally used. A point of view as stated above, must assume that there is innate knowledge of the possibilities of the vocal tract and of its relations with the (initially empty) perceptual representation.

Below, a number of investigations will be mentioned in which the relation between perceptual development and speech acquisition is studied. Snow (1964) showed in her study that speech errors made by children resemble very much the perceptual confusions of speech sounds by adults as reported by Miller and Nicely (1965). Olmsted (1966) developed a theory of speech acquisition based on the mentioned study by Miller and Nicely. This 'perceptual' theory predicts that the speech errors made by the child, as well as the order in which speech sounds are mastered, is determined by the perceptual discriminability of the speech sounds. It is remarkable that Olmsted does not mention Snow's article, neither in his 1966 paper nor in his book of 1971, in which he tested his theory with the help of acquisition data. From this latter investigation it follows that the prediction with respect to the occurrence of speech errors is affirmed, the prediction concerning the developmental course, however, appears to be incorrect.

Winitz (1969) who reviewed a large number of studies that examined the relation between auditory discrimination and articulatory performance, finally concludes that '...the evidence overwhelmingly supports the point of view that articulatory defective children score below non-articulatory defective children on tests of speech sound discrimination.' (p. 185).

It does not follow from this conclusion, however, that motor development reflects perceptual development. Indeed, Locke (1972) reports a correlation of .07 between correct articulation and correct perception of consonants by 3-year old children. This latter finding very clearly indicates that a simple 'perceptual' theory is not sufficient: articulatory performance cannot be predicted merely by perceptual development, at any

rate not when it is measured thus.

Finally, we report a study that actually determined the effect of perceptual training on articulatory acquisition. Winitz and Preisler (1965) found that at least some children that participated in the experiment improved their articulation of speech sounds if they had been given sound-discrimination training beforehand. It should be noted, however, that not all subjects showed this improvement and that the control group also showed improvement. Moreover, the possibility is not excluded that, during the perceptual training, also articulation training occurred, since in the perceptual training, words were used.

The cited findings indicate that a purely perceptual theory of speech acquisition cannot be acceptable. They do, however, indicate the existence of a dependency and stress the priority of the perceptual development.

The two abovementioned phenomena, which served as an argument for the hypothesis that perception is the only determinant of articulatory behaviour, still remain difficult to explain. They show at any rate that the learning system can infer new relations between articulatory behaviour and perceptual effects from a restricted experience.

We might summarize the foregoing discussion by proposing that under normal circumstances speech acquisition is characterized by both a perceptual and a motor development. Perceptual development must be supposed to preceed motor development since this provides the norm to which development must conform. The relation between both domains is explored by means of a matching procedure which is not random, but which reflects the properties of the strategies followed by the subject. In certain stages of the development these strategies find expression in considerable discrepancies between input and output of the speech generating system.

2.2. Possibilities of speech correction for the deaf

The method that is followed by the hearing child during speech acquisition, of which we have analysed a number of aspects in the previous section, cannot be used by the deaf. At any rate, not in the natural situation where there are no possibilities for the deaf to receive the speech product on which matching procedures could be applied. Nowadays, there is the possibility of greatly amplifying the acoustic signal by means of specially adapted hearing aids which do provide the deaf with some acoustic input; it does not solve the problem, however.

Besides, one could propose that the deaf can make use of visible articulatory actions of speakers in his surroundings. Even if we were to assume that the deaf were able to match these movements (he cannot see his own movements) one should still consider that these movements contain only limited information on the speech product as well as on the articulatory processes from which it results.

In this context, it must be stressed that the only complete and intended speech product is the acoustic signal; therefore we believe that the acoustic signal only contains the adequate norm that can guide the development of speaking. In this view, the visible articulatory movements are considered an accompanying correlate of the process of speaking, which, for that reason, cannot form an acceptable basis for speech learning.

The only method of learning speech that remains for the deaf, is the method discussed above which was named 'Learning through process shaping'. Essential to the method is its motor approach, by which is meant that the skill is taught by supplying the pupil with a detailed description of the motor aspects of the desired behaviour. Two differences must be mentioned between the pupil that learns to play the flute and the deaf learning to speak, both using this method. Firstly, the deaf pupil does not receive auditory feedback, while the pupil that practises flute playing does. Secondly, in contrast to flute playing, speaking does not belong to the type of skills that consist of a relatively small repertoire of actions which can be easily specified. On the contrary, as was shown in the first chapter, the specification of relevant actions during speaking still form a real problem.

Below, we shall mention a number of difficulties that are related to the 'Learning through process shaping' method of speech training, which is the method that was and still is used today in the speech education of the deaf.

1. The first difficulty is related to the problem of invariance and the presentation of the norm. In the first chapter it was shown that no invariant relationship exists between perceptual and production (motor) units, thus it could be concluded that different articulatory positions (with correspondingly different neural commands) can be used for the realization of one perceptual unit (phoneme). It was also demonstrated that distortions of the vocal tract can be compensated with a modified adjustment of the articulators. This again proves that perceptually identical speech sounds can be produced with different oral configurations. Other practical examples of phonemes that can be produced by means of different articulatory positions are the /l/ which can be produced both unilaterally and bilaterally, the /v/ and the /w/ which can be produced labiodentally and bilabially, the /d/ and the /t/ which can be produced alveolarly and also with the tip of the tongue against the lower teeth. Presumably most of the differences will be considerably smaller than the ones mentioned here.

A consequence of this is that a pupil, taught via the motor approach, will have great difficulty in procuring and internalising norms for the various speech sounds. Moreover, for each specific phonetic unit he will be required to learn one standard articulatory configuration. However, the speech trainer possesses a perceptual norm* so that he will use his ear when evaluating whether a speech sound or sequence is correctly produced. The deaf pupil on the other hand now notices that more than one motor realization is passed as correct by the teacher. The result of this is that the deaf pupil is confronted with a very tricky pattern recognition problem, for which a solution on these lines (i.e. without really supplying a norm) is perhaps impossible.

2. A second problem is related to the specifiability of the actions involved in the process of speaking. The specification of the movements of the different articulators during speaking is undoubtedly a very complicated matter. Lenneberg (1967) has determined that some hundred differ-

* *We shall not here enter into the problematical matter of shifts in criterion which will inadvertently affect the speech trainer when he has to judge repeated speech sounds.*

ent muscles are involved in speaking including muscles of the chest, stomach, neck, face, pharynx, larynx and oral cavity. Taking an average speaking rate of 14 phonemes per second, he assumes that 14 times per second a neural command is sent to each of the muscles involved; moreover, these commands have to be very precisely coordinated. It will be clear that a specification of the necessary articulatory movements for a speech act in terms of commands to the separate muscles is not only impossible but also quite undesirable since such an instruction has no psychological significance. Nobody can be expected to have the capacity of contracting a specific muscle to a certain degree, and certainly not of realizing an instruction which states to what extent each of a set of muscles must be contracted or decontracted and in what order precisely. The smallest functional unit of action is not the contraction or decontraction of some muscle, but a movement of some part of the body.

In practice, the speech trainer will make use of descriptions or indications concerning the behaviour of the articulators as traditionally employed in phonetics. These instructions mainly concern the static aspects of this behaviour. They may be sufficiently adequate when they concern states and behaviour of those structures of the vocal tract that can be easily pointed out, much less, however, when they concern less peripheral structures and presumably even less adequate when they concern the dynamic aspects of the behaviour and the way in which the different movements must be coordinated. Working in this way the speech trainer will attempt to transfer the 'Plan' of actions of the skill to the pupil. Miller, Galanter and Pribram (1960) have studied the question whether this approach can be considered fruitful in teaching a skill. They come to the conclusion that the terms in which skilled behaviour is generally described must be considered inadequate: the description does not actually help the pupil to execute the desired behaviour since it does not actually appeal to motor dimensions. They state that: '... it may actually be better pedagogy to let the student invent his own idiosyncratic tactics for carrying the Plan into his muscles.' (p. 83). This certainly is a very pessimistic point of view with respect to the possibilities of instruction in skill acquisition. It implies that the only thing that can be done is to show the pupil the goal at which he has to aim. Since this method is obviously quite impossible with the deaf, the teacher of the deaf is obliged to resort to instructions that try to describe as far as possible the 'Plan' of actions.

Paradoxally as it may seem, instructions for eliciting some behaviour

are often only effective as far as they can be stated in terms of some effect that results from it. Thus, in order to get somebody to breathe with his diaphragm, it can be very helpful to let him imagine that he inhales the odour of a flower, the physician that wishes to look into a child's throat will ask him 'Say, ah'. These phenomena mean that often behaviour can only be evoked via the perceivable result of it. Remember that this is characteristic for behaviour that is mastered via 'output matching'. From the articulatory behaviour, only the visible external articulatory movements can more or less be phrased in such perceptual terms. Indeed these movements cause but a minor problem for the deaf.

3. A third difficulty is connected with the fact that the verbal utterances of the deaf pupil are always judged by an external authority, namely the speech trainer who, in giving feedback, normally will not be able to give attention to more than one aspect at a time. Often he will only say whether a speech sound was pronounced correctly or not. If the sound is incorrectly pronounced (according to the momentary criterion of the speech trainer) the pupil is invited to pronounce the sound again, after which he is given additional instruction with respect to some motor aspect. Actually, the pupil misses much information from his incorrect attempts since he does not know whether this speech sound satisfied the criterion in other respects, or in what way it departed from this criterion (the sound produced may very well closely resemble another speech sound). This is a consequence of the fact that information is always given in motor terms and not in terms of deviation from the norm itself.

The objections mentioned all have a theoretical character. In practice, the method leads to positive results, especially when the difficulties mentioned are taken into account. This does not mean, however, that the speech of even the most perfectly trained deaf could not be improved intelligibility and naturalness still leave much to be desired.

The difficulties connected with the motor approach can in principle only be relieved by giving the deaf that information that contains the only valid norm the speech product This information must be 'perceptual' rather than physical. In other words, it is not sufficient to display the physical signal directly in some way, it must be 'processed' in the same way as occurs in human auditory perception. Through this processing, the physical information is transformed into perceptual information, which can form a sound basis for matching procedures on the part of the pupil.

Because of the very serious problems that are involved in attempting to supply information of this kind to the deaf, such an attempt should not be considered an alternative to the traditional method, but rather a valuable aid to be integrated in it.

In the two papers that follow, the development and evaluation is described of a device that can supply information on the speech product.

Psychol Res 37, 51–70 (1974)

Development of a Vowel Corrector for the Deaf

Dirk-Jan Povel

Department of Psychology, University of Nijmegen, Nijmegen, The Netherlands

Received January 29, 1974

Summary The development of a visual speech apparatus that gives information about the identity of vowels spoken in a CVC context is described. The principle of the apparatus is based on a dimensional analysis of the vowel spectra. In developing the device, the demands of the specific speech correction setting of the deaf were taken into account as much as possible. From a number of physical and statistical analyses of vowels in different consonantal surroundings it appeared possible to separate two vowels quite satisfactorily. In describing the apparatus, special attention is given to some technical features that were added to satisfy the requirements of speech correction. Finally, the operation of the device is tested with a large number of monosyllables pronounced by 20 speakers. The results indicate that the apparatus satisfies the conditions stated for practical use.

In their 1967 article Plomp, Pols and Van de Geer described a technique based on a dimensional analysis of vowel spectra, for representing vowels within a limited number of dimensions. A two-dimensional representation can be made visible as a light spot on an oscilloscope screen. Different vowels are projected on different loci of the screen. The authors suggested that this technique 'might have some value as a visual feedback system for speech training of the deaf'. The present article reports the steps that were taken from this suggestion to the construction of a practicable articulation corrector for the deaf which was ultimately called 'Vowel Corrector'. Furthermore, an experiment is described which was designed to evaluate the working of the apparatus.

Construction of the Vowel Corrector

One can divide visible speech apparatus into four types: pitch and intonation correctors, intensity correctors, rhythm correctors and articulation correctors. Each type has its special requirements. Articulation correctors, to which class our vowel corrector belongs, have, as a class, a comparatively great number of requirements among which several are hard to realize technically. Therefore, it seems useful to first make explicit the most important requirements for an articulation

corrector before proceeding to the description of their implementation in the construction of the Vowel Corrector.

Requirements for an Articulation Corrector

1. An articulation corrector should display in a unique way the different speech sounds for which it is designed. This means that use must be made of those aspects of the physical information which yield representations differentiating between different speech sounds. It is surprising that uniqueness of representation has been given so little attention in existing articulation correctors. Only in recent years has there been some interest on this point. Reich and Weed (1972) computed communality measures of the physical displays of different sounds on the "visual vocoder", a device that displays spectral information. The physical discrimination they found is far from perfect. Besides, the only criterion that is of practical importance is the discrimination by subjects. This was studied by Schulte (1971) on a device that represents sounds as Lissajous figures: he determined the discriminability of specific features of speech sounds from different displays. The results were disappointing, which is illustrated by the phenomenon that the discrimination appeared to be dependent on age and intelligence of the subjects.

2. Displays produced by an articulation corrector must be easily interpretable. A device may give unique displays for different sounds but when it takes time for the subject to decide what the display really means, the information loses its values since the feedback process is interrupted. This will generally be the case with articulation correctors that yield complex representations as for instance devices that display the spectral information directly (Potter *et al.*, 1947; Searson, 1965; Risberg, 1968; Stark *et al.*, 1968, 1970; Nordman, 1972; Nickerson and Stevens, 1972; Kisner and Weed, 1972) and devices that display Lissajous figures (Pronovost *et al.*, 1967, 1968; Montgomery, 1970; Schulte, 1971). If one wishes to avoid such an interpretation process on the part of the subject, the dimensionality of the display should be reduced considerably, with maintenance of the informational content, however. A number of devices have been designed for the unambiguous display of spectral information, for instance by means of light spots on an oscilloscope screen. Examples are the ADL sustained phoneme analyzer (Cohen, 1968), The Gallaudet visible speech trainer (Pickett and Constam, 1968), The A.P.I. (Kalikow and Klatt, 1970, 1972) and the Three Parameter Display of Ferber and Weed (1972). Our Vowel Corrector also belongs to this category. It is regrettable that none of the just mentioned authors has given a quantitative measure of the discriminative power of their devices. The more so because we had reason to doubt whether any general speech-sound display system will ever reach the resolving power which is required for

effective practical use. This doubt was in part based on research of, for instance, Pols *et al.* (1973) who clearly showed that a satisfactory separation of all vowels along two dimensions is hardly possible. For our case it appeared necessary, even with the use of a very powerful separation procedure, to restrict the number of vowels to be displayed.

3. An articulation corrector should display speech sounds which are produced in normal articulatory context. From speech acquisition theory and practice it is known that training of isolated speech sounds only is insufficient. In terms of McNeilage (1970): knowing the goal position is only the basis besides, a speaker must find all roads i.e. motor commands to go from all possible initial positions to a certain goal position. The deaf child, therefore must be given an opportunity to practice the pronunciation of a speech sound in different contexts, in units at least as large as a syllable. It is therefore regrettable that most articulation correctors can only clearly display isolated pronounced sounds. There are however some exceptions. In sound spectrograms that portray intensity fluctuations of the different frequency components as a function of time, a specific phoneme in a word can be pointed out (Stark *et al.*, 1968, 1970; Kisner, 1972). The Lucia spectrum indicator (Risberg, 1968) has a possibility to "freeze" a sound by pushing a button while speaking a word, but we believe that this method of segmentation will not be feasible in practice. With the S-indicator (Risberg, 1968; Martony, 1969) the pronunciation of the S-sound can be practised in all contexts, since the S is easily segmentable due to the very specific characteristics of this sound.

To realize the mentioned requirement, our Vowel Corrector is equipped with an automatic segmentation device so that only the sound to be trained can be displayed. Moreover we took care that the different allophones of a speech sound are displayed by the apparatus in one cluster.

4. The interval between enunciation of the speech sound and the feedback from the articulation corrector should be short. Because the given feedback is of the learning-feedback type (Annett, 1969) which means that the feedback reaches the subject after the response has been made, it is important to make the delay between response and feedback as short as possible. Only then can the subject relate the augmented feedback with the intrinsic feedback corresponding to the response, which gives the subject the opportunity to learn which intrinsic feedback belongs to a right response. In our case the automatic segmentation procedure introduces some delay, but this could be kept very short.

5. The apparatus should work reliably and should be easy to operate. Also our decision not to construct a computer-aided system, but to restrict to hardware design was determined by a practical requirement. Though we realized that the use of a computer could have a number of

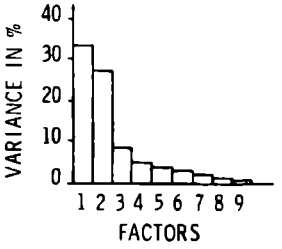

Fig 1 Percentages of the total variance explained by the first nine factors

advantages (cf Nickerson and Stevens 1972) the most important being the flexibility in designing the system and the possibility of adjusting to special demands our choice of constructing a hardware apparatus was deliberate since the Vowel Corrector has ultimately to be used in an Institute for the Deaf, where no computer is available

Plomp's Factor Display

The automatic vowel recognition system developed by Plomp, Pols and Van de Geer (Plomp *et al* 1967) is not based on formant extraction as many traditional systems are, but starts from all information available in the spectrum The spectrum is divided into a number (N) of frequency bands Each vowel spectrum is represented by the energy level on these frequency bands In this way all vowels can be described in an N dimensional space By means of a principal component analysis it appeared to be possible to reduce the number of dimensions without losing much information This method yields a number of new orthogonal dimensions, called factors, which are linear combinations of the original ones Each of the new factors accounts for a proportion of the total variance see Fig 1 This Figure is based on an article by Klein, Plomp and Pols (Klein *et al*, 1970) which is an extension of the above mentioned 1967 article Fig 1 gives the variance contributions of 9 new dimensions computed from a total of 600 vowels, being 12 different vowels each pronounced by 50 speakers in an *h t* context (Klein 1970) The first two factors are by far the most important ones explaining 33 7% and 27 2% of the total variance respectively Together this is 60 9% of the variance When more dimensions are added, the amount of explained variance increases slowly with 3 factors 69 6%, with 4 factors 75 4% The authors determined the success of an automatic vowel recognition using a varying number of dimensions The identification proceeded on a maximum likelihood basis Identification scores were determined with an without speaker normalization This normalization was achieved by translation

Table 1. Identification scores in % found by Klein *et al.* (1970), with and without speaker-dependent correction, using various numbers of factors

	Number of factors used				
	1	2	3	4	6
Original data	51.0	78.2	86.7	88.7	93.2
Data corrected by translation	60.2	88.0	97.2	97.5	97.7

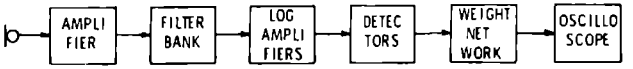

Fig. 2. Block diagram of the "Factor Display"

of the individual centres to their mean. In Table 1 the identification scores are given with use of one, two, three, four and six factors, with and without speaker dependent correction.

In the 1967 paper a device was described — later called the "Factor Display" — that displayed the position of a vowel in a plane composed by the first two factors. In this way each vowel had its own projection region. A block diagram of this apparatus is given in Fig. 2. The processing can be summarized using the symbols X_i ($i = 1, 2, \ldots N$) for the Filter outputs in dB, w_{1i} and w_{2i} for the two sets of weightings and L for the overall sound pressure level (SPL). The coordinates of the spot on the screen are specified as: $F_1 = \sum_{i=1}^{N} (L - X_i) \cdot w_{1i}$ and $F_2 = \sum_{i=1}^{N} (L - X_i) \cdot w_{2i}$. This processing entails a loudness correction.

This "Factor Display" was thoroughly tested by the investigator in order to determine its feasibility as an articulation corrector. Then it appeared that the system lacked some of the properties that were considered indispensable for an articulation corrector. First of all, the identification score i.e. the discriminative power based on the first two factors is rather low for our purpose. Klein *et al.* (1970) report an identification score of 88% with speaker normalization and 78% without, see Table 1. Evaluating these data, one should realize that these are identification scores for 12 vowels pronounced by male speakers in only one consonantal context (*h* — *t*). When used as an articulation corrector, however, ideally all allophones should be identified correctly. But the identification score will supposedly decrease by introducing other allophones, since coarticulation has a large effect on the spectral properties of vowels. This is exactly what we found when large numbers of

vowels out of several CVCs were displayed on the Factor Display increasing the number of allophones increased the overlap of projection regions of different vowels considerably Moreover it was noticed while working with this apparatus that an imperfect loudness correction can destroy the discrimination another condition for a reliable display is that the input signal remains between an upper and a lower limit

On the basis of our experience in testing the system we decided to confine ourselves to only two vowels in order to make the apparatus suited for actual speech correction Moreover a number of technical modifications had to be carried out the most important being the construction of an automatic segmentation device These technical aspects will be discussed later There are two reasons why a choice was made for the vowels /*I*/ and /*ɛ*/ First of all, deaf children appear to have special problems pronouncing these vowels Secondly, these two dutch vowels were chosen because they are physically very similar in the formant plane for example they are neighbours (Pols *et al* , 1973) if we were to succeed in satisfactorily separating these two vowels, then we should certainly be able to separate most other pairs of vowels

The restriction to two vowels implies that a ' discriminant analysis ' is now the most appropriate separation technique for determining optimal weightings for the different band filters The principle of discriminant analysis as applied to the present problem will be explained below Therefore, discriminant analyses were computed for /*I*/ and /*ɛ*/ sounds from different consonantal contexts and produced by male as well as by female voices From these analyses, which will be discussed below it had to be decided whether an acceptable separation of /*I*/ and /*ɛ*/ sounds could be realized and whether this separation is speaker dependent

Analyses and Computations on /I/ and /ɛ/ Sounds

A list consisting of 54 monosyllabic words was pronounced twice by 4 male and 4 female speakers and recorded on tape One half of the words contained the vowel /*I*/, the other half the vowel /*ɛ*/ For the construction of the 48 CVCs in the list, 12 different initial and 6 different final consonants were used Also four VCs and the isolated vowels were included Thus 864 words were collected, half of them with vowel /*I*/ and half with vowel /*ɛ*/ From each of the recorded words a segment was singled out from the steady part of the vowel A spectral analysis of these segments resulted in 17 values for each of the 864 vowels, namely the SPLs on 17 frequency bands covering the frequency range from 125 Hz to 8,000 Hz The centre frequencies of the 17 frequency bands are given in Table 2 except for the first two filters all are 1/3 octave

Next, two discriminant analyses were executed upon these data Given two groups measured on N variables discriminant analysis

Table 2. Centre frequencies of the band-pass filters used

Filter Number	1	2	3	4	5	6	7	8	9
Center freq. in Hz	125	225	315	400	500	630	800	1000	1250
Filter Number	10	11	12	13	14	15	16	17	
Center freq. in Hz	1600	2000	2500	3200	4000	5000	6400	8000	

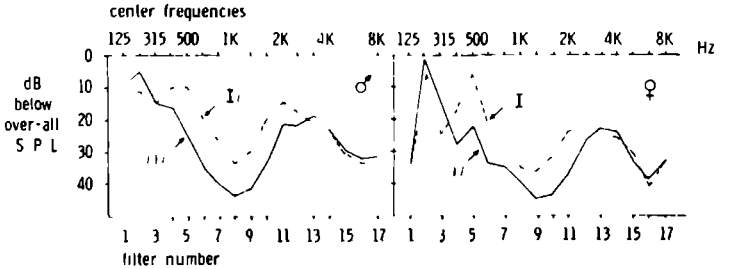

Fig. 3. Average frequency spectra of /I/ and /ɑ/. Left male, right female speakers

determines a linear combination of the original variables in such a way that the between-group variance is maximized as compared with the within-group variance. The discriminant dimension can therefore be specified as

$$Z = w_1 X_1 + w_2 X_2 + \cdots + w_N X_N ,$$

where X_i is the ith variable, and w_i its weight. (For details of this procedure see Cattell, 1966).

One discriminant analysis was done on the male data, one on the female data. The results of both analyses will be discussed jointly in order to facilitate comparison. A precise indication of the discriminative capacity of the solution is the value of η^2, which is the ratio of the between-group variance and the total variance of the scores on the discriminant dimension. The maximum value of η^2 is 1, the higher η^2 the better the solution. For the male vowels a solution was found with an $\eta^2 = .917$, for the female vowels $\eta^2 = .894$. Some further aspects of these analyses are presented in Figs. 3–7. Fig. 3 presents the means of /I/ and /ɑ/ on the 17 frequency bands for the male and female speakers, thus forming the average frequency spectra.

Fig. 4 shows the total variance and the variance explained by the differences between vowels per filter.

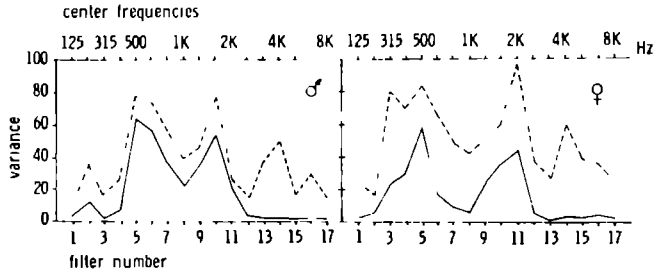

Fig. 4. Total variance (dashed line) and between-vowel variance (continuous line) per filter. Left male, right female speakers

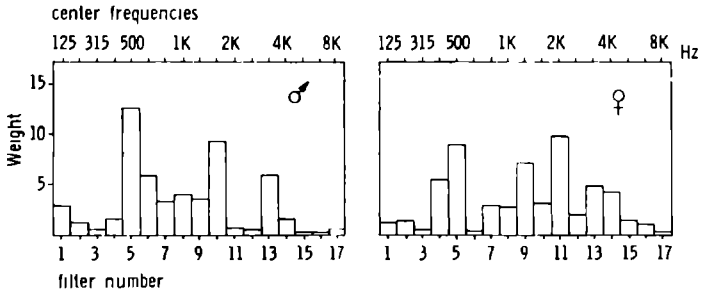

Fig. 5. Standard weights for the 17 filters. Left male, right female speakers

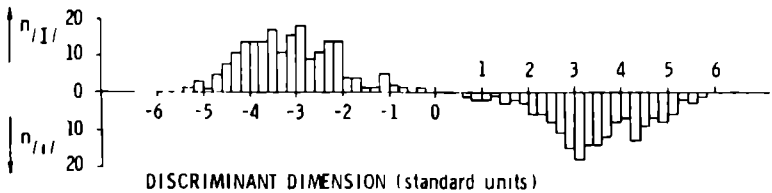

Fig. 6. Distribution of /I/ and /i/ along discriminant dimension, male vowels $N_{/I/}$ 197; $N_{/i/}$ 182

Fig. 5 presents the absolute values of the weights for the 17 dimensions as computed for the male and female data separately. All weights are given in arbitrary units of variance for the particular dimension.

Figs 6 and 7 give distributions on the discriminant dimension, Z, for /I/ and /i/. Fig. 6 for male and Fig. 7 for female speakers As could

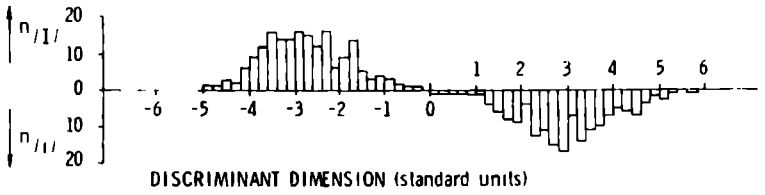

Fig. 7. Distribution of /I/ and /i/ along discriminant dimension; female vowels. $N_{/I/}$ 182; $N_{/i/}$ 169

be expected from the high values of η^2, the plots of the discriminant scores show a complete separation of /I/ from /i/, there is no overlap, and even a considerable gap.

The discrimination of the female vowels is not as good as that of the male vowels. This is not surprising when the percentages of between-vowel variance of the male and female data in Fig. 4 are compared. We must be careful not to generalize these findings too soon, since relatively small samples were analyzed. However, it is obvious from the analysis that the spectral characteristics of the male and female vowels do show large differences. Therefore it did not seem reasonable to perform a discriminant analysis on the pooled data of men and women together. Later such an analysis will be reported for a data set which was reduced to 5 dimensions.

At this point, the question arose whether it was really necessary to use all 17 dimensions for the discrimination. Klein *et al* (1970) did in fact demonstrate that even a considerable reduction of the number of frequency bands only has a slight negative effect on the automatic recognition of vowels. Therefore, and for economical reasons, the possibility of constructing an apparatus with only 5 input channels was examined. In order to test the feasibility of this reduction to 5 input channels, the discriminant capacity was determined for a number of combinations of five frequency bands, chosen on the basis of the outcomes of the former analyses. Combinations of 1/3-octave and of 2/3-octave filters were tested. The η^2 and distance between centroids of /I/ and /i/ for the different solutions are given in Table 3. For the sake of comparison the original solutions based on 17 dimensions are also presented in this table.

All solutions for five filters are quite acceptable as far as discrimination is concerned, although none of them is as good as the solution based on all 17 filters. It needs mention that even for the pooled vowels (bottom row Table 3) a very reasonable discrimination is found. It was therefore decided to proceed with the construction of a five input channel device.

Table 3. The results of the analyses using different combinations of filters in terms of η^2 and the distances between the centroids. For center frequencies corresponding with the filter numbers see Table 2

Number of filters	Speakers	η^2	distance centroids
17 1/3-octave	male	.917	6.63
17 1/3-octave	female	.894	5.80
5 1/3-octave			
5, 6, 8, 10, 13	male	.903	6.09
5, 6, 7, 9, 10	male	.889	5.65
4, 5, 9, 11, 13	female	.874	5.28
4, 5, 9, 10, 11	female	.794	3.92
5 2/3-octave			
400 3200 Hz	male	.855	4.85
400 3200 Hz	female	.884	5.51
500 4000 Hz	male	.890	5.70
500 4000 Hz	female	.853	4.80
	male + female	.843	5.63

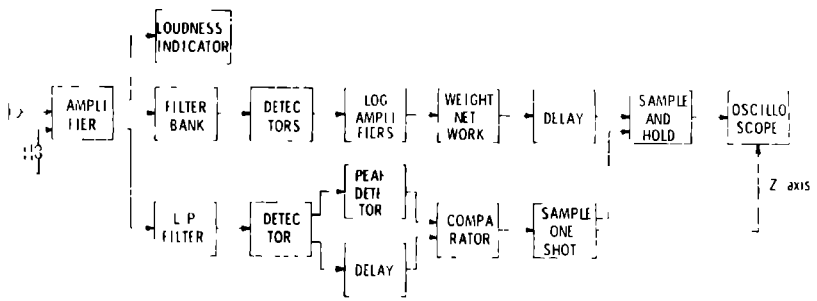

Fig. 8. Block diagram of the Vowel Corrector

Design of the Vowel Corrector

Fig. 8 and 9 present a block diagram and a photograph of the Vowel Corrector in its final form, respectively. The system consists of two parts with quite different functions.

The upper branch is the core of the device: here the processing of the signal takes place. It is essentially the same as Plomp *et al*'s Factor Display: the only additions are a Sample and Hold system and a delay line. The Sample and Hold system samples and integrates the incoming signal during a predetermined period of time. At the end of the interval the integrated value of the sample is fed as a DC level to the Y-axis of the

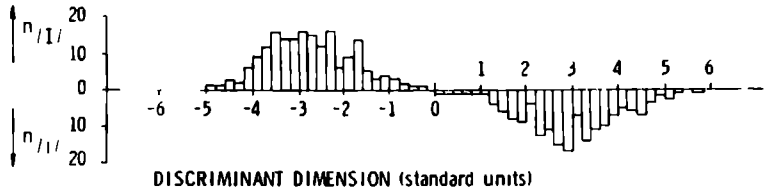

Fig. 7. Distribution of /I/ and /ɛ/ along discriminant dimension, female vowels. $N_{/I/}$ = 182, $N_{/ɛ/}$ = 169.

be expected from the high values of η^2, the plots of the discriminant scores show a complete separation of /I/ from /ɛ/: there is no overlap, and even a considerable gap.

The discrimination of the female vowels is not as good as that of the male vowels. This is not surprising when the percentages of between-vowel variance of the male and female data in Fig. 4 are compared. We must be careful not to generalize these findings too soon, since relatively small samples were analyzed. However, it is obvious from the analysis that the spectral characteristics of the male and female vowels do show large differences. Therefore it did not seem reasonable to perform a discriminant analysis on the pooled data of men and women together. Later such an analysis will be reported for a data set which was reduced to 5 dimensions.

At this point the question arose whether it was really necessary to use all 17 dimensions for the discrimination. Klein *et al.* (1970) did in fact demonstrate that even a considerable reduction of the number of frequency bands only has a slight negative effect on the automatic recognition of vowels. Therefore, and for economical reasons, the possibility of constructing an apparatus with only 5 input channels was examined. In order to test the feasibility of this reduction to 5 input channels the discriminant capacity was determined for a number of combinations of five frequency bands chosen on the basis of the outcomes of the former analyses. Combinations of 1/3-octave and of 2/3 octave filters were tested. The η^2 and distance between centroids of /I/ and /ɛ/ for the different solutions are given in Table 3. For the sake of comparison the original solutions based on 17 dimensions are also presented in this table.

All solutions for five filters are quite acceptable as far as discrimination is concerned, although none of them is as good as the solution based on all 17 filters. It needs mention that even for the pooled vowels (bottom row Table 3) a very reasonable discrimination is found. It was therefore decided to proceed with the construction of a five input channel device.

Table 3. The results of the analyses using different combinations of filters in terms of η^2 and the distances between the centroids. For center frequencies corresponding with the filter numbers see Table 2

Number of filters	Speakers	η^2	distance centroids
17 1/3-octave	male	.917	6.63
17 1/3-octave	female	.894	5.80
5 1/3-octave			
5, 6, 8, 10, 13	male	.903	6.09
5, 6, 7, 9, 10	male	.889	5.65
4, 5, 9, 11, 13	female	.874	5.28
4, 5, 9, 10, 11	female	.794	3.92
5 2/3-octave			
400 3200 Hz	male	.855	4.85
400 3200 Hz	female	.884	5.51
500 4000 Hz	male	.890	5.70
500 4000 Hz	female	.853	4.80
	male + female	.843	5.63

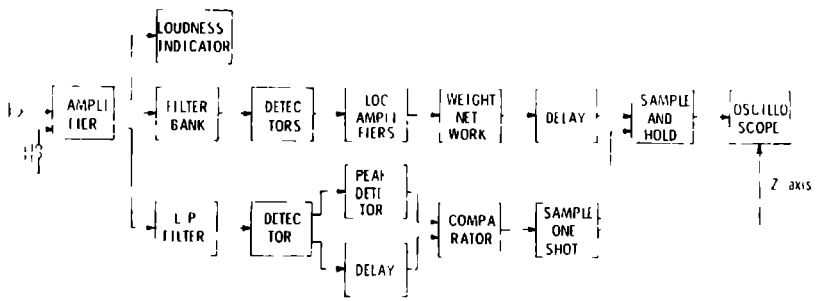

Fig. 8. Block diagram of the Vowel Corrector

Design of the Vowel Corrector

Fig. 8 and 9 present a block diagram and a photograph of the Vowel Corrector in its final form, respectively. The system consists of two parts with quite different functions.

The upper branch is the core of the device; here the processing of the signal takes place. It is essentially the same as Plomp *et al*'s Factor Display: the only additions are a Sample and Hold system and a delay line. The Sample and Hold system samples and integrates the incoming signal during a predetermined period of time. At the end of the interval the integrated value of the sample is fed as a DC level to the Y-axis of the

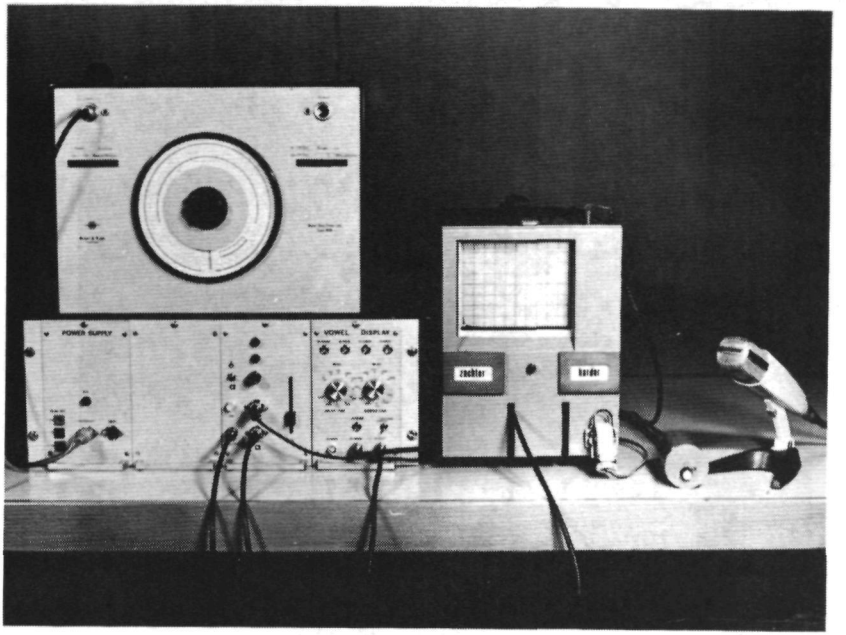

Fig. 9. The Vowel Corrector

oscilloscope (Tektronix 5103N); simultaneously the intensity suppression is temporarily removed, allowing the signal to be seen. Since the input of the X-amplifier is grounded, the level is displayed as a spot, and as the oscilloscope is in storage mode the displayed spot remains on the screen as long as required. The total time delay between the end of the pronounced word and the display of the spot is on the average 250 msec., which for our practical purpose, is a negligible value.

The lower branch fulfills a segmentation and timing function: it takes care that the result of the processing of only the vowel section of the word is made visible. The automatic segmentation is based on the observed phenomenon that the intensity of vowels is higher than the intensity of consonants. This holds in most cases even for the /*i*/ which is the softer of the two vowels. For some speakers this relation does not hold when the consonants in the word are formed by voiced consonants or fricatives. This problem could be solved by adding a 3rd order Low Pass Filter (3 dB point at 2.7 kHz). The segmentation is carried out as follows: the incoming signal is both fed into a peak detector and a delay line. In the peak detector the peak of the signal, i.e. a CVC, is determined during a period of 420 msec. which is ample time for the pronunciation of a CVC.

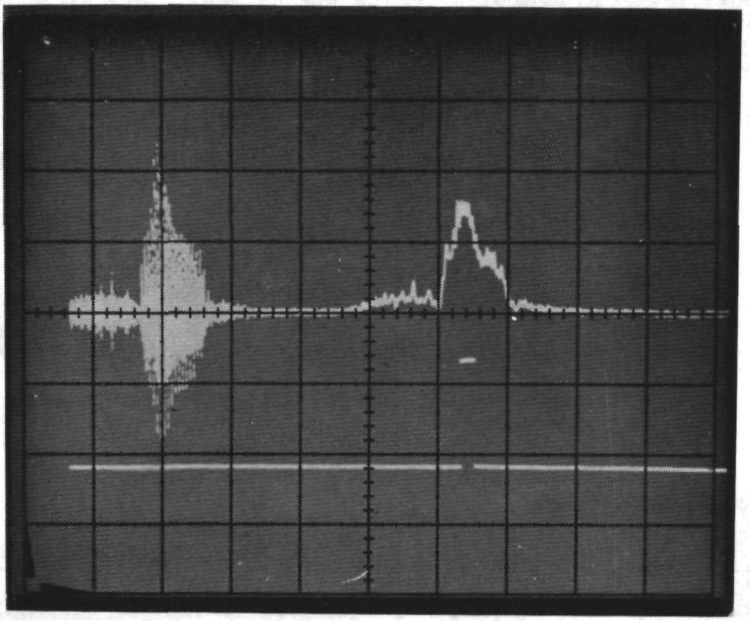

Fig. 10. Segmentation of a vowel section from a CVC (see text)

This time period starts when the speaker has pressed a button and when the signal passes a fixed intensity value (in practice a little higher than the background noise). This duration is just as long as the delay time. The peak value is fed to a comparator which, 420 msec. later, receives the delayed signal. At the moment that the peak level and the envelope of the delayed signal are the same, a pulse is given to a one shot. This one shot determines starting and stopping point of the Sample and Hold circuit. Now it may be clear why also in the upper branch a delay line is included. Since both delay lines have exactly the same delay, the Sample and Hold circuit samples precisely in the central part of the vowel of the CVC. Sample duration is 20 msec. which means that a segment of only 20 msec. out of the vowel is displayed. This may seem rather short. Making the sample time longer, however, increases the chance of segmenting in one of the adjacent consonants. Moreover no significant differences in the display of vowels were found with different segment lengths, provided that the segment was in the vowel part. In Fig. 10, a photograph of the segmentation is shown. The upper line presents the incoming signal */sIp/* and the delayed signal which is envelope detected; the lower line shows the gate which opens during the passage of the central portion of the

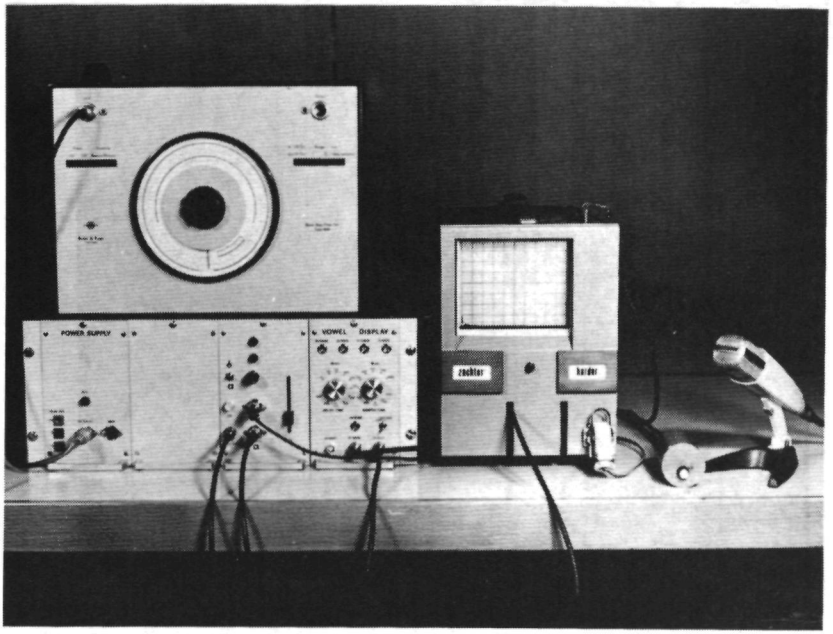

Fig. 9. The Vowel Corrector

oscilloscope (Tektronix 5103N); simultaneously the intensity suppression is temporarily removed, allowing the signal to be seen. Since the input of the X-amplifier is grounded, the level is displayed as a spot, and as the oscilloscope is in storage mode the displayed spot remains on the screen as long as required. The total time delay between the end of the pronounced word and the display of the spot is on the average 250 msec., which for our practical purpose, is a negligible value.

The lower branch fulfills a segmentation and timing function: it takes care that the result of the processing of only the vowel section of the word is made visible. The automatic segmentation is based on the observed phenomenon that the intensity of vowels is higher than the intensity of consonants. This holds in most cases even for the /*i*/ which is the softer of the two vowels. For some speakers this relation does not hold when the consonants in the word are formed by voiced consonants or fricatives. This problem could be solved by adding a 3rd order Low Pass Filter (3 dB point at 2.7 kHz). The segmentation is carried out as follows: the incoming signal is both fed into a peak detector and a delay line. In the peak detector the peak of the signal, i.e. a CVC, is determined during a period of 420 msec. which is ample time for the pronunciation of a CVC.

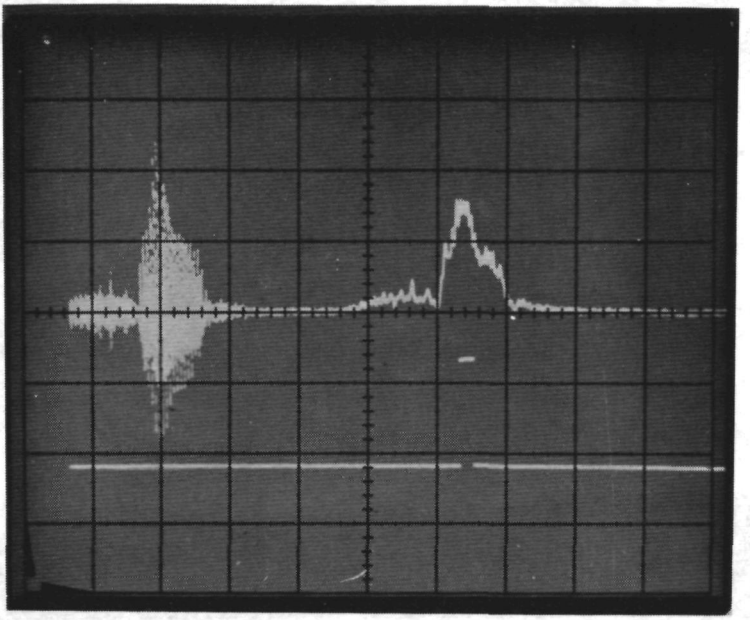

Fig. 10. Segmentation of a vowel section from a CVC (see text)

This time period starts when the speaker has pressed a button and when the signal passes a fixed intensity value (in practice a little higher than the background noise). This duration is just as long as the delay time. The peak value is fed to a comparator which, 420 msec. later, receives the delayed signal. At the moment that the peak level and the envelope of the delayed signal are the same, a pulse is given to a one shot. This one shot determines starting and stopping point of the Sample and Hold circuit. Now it may be clear why also in the upper branch a delay line is included. Since both delay lines have exactly the same delay, the Sample and Hold circuit samples precisely in the central part of the vowel of the CVC. Sample duration is 20 msec. which means that a segment of only 20 msec. out of the vowel is displayed. This may seem rather short. Making the sample time longer, however, increases the chance of segmenting in one of the adjacent consonants. Moreover no significant differences in the display of vowels were found with different segment lengths, provided that the segment was in the vowel part. In Fig. 10, a photograph of the segmentation is shown. The upper line presents the incoming signal */sIp/* and the delayed signal which is envelope detected; the lower line shows the gate which opens during the passage of the central portion of the

vowel. The delay is realised with a so-called "Bucket brigade Delay Line" (Philips TCA590) ref. Sangster (1970), which is a shift register for analog signals consisting of 512 condensors. Using a clock-frequency of 610 Hz a delay of 420 msec. is obtained.

It was noticed earlier that the intensity of the input signal should remain between an upper and a lower limit. Because it is known that, especially for the deaf, it is difficult to speak on a more or less constant level, a loudness indicator was built. This device checks during the speaking period of 420 msec. whether the signal intensity exceeds a fixed level which would mean overload of the system. When this happens a panel labeled "softer" is lighted. When during the speaking period the signal never reaches a predetermined lower level a panel "louder" is lighted. In both cases no display will appear on the oscilloscope. As it was decided to evaluate the Vowel Corrector on male deaf pupils, the weights associated with the two best solutions for male voices (see Table 3) were tested separately. It then appeared that the display based on 5 1/3-octave filters had one great disadvantage: many other vowels were projected on the same region of the discriminant dimension as the /*I*/ and /*i*/ sounds. This was not the case with the discriminant dimension based on 5 2/3-octave filters in the frequency range from 500 to 4000 Hz: here all other vowels (except the /*y*/) appeared to project outside the /*I*/-/*i*/ region. The /*y*/ is on the average, positioned somewhere between the /*I*/ and /*i*/. So it was decided to use the separation procedure based on the five 2/3-octave filters in the frequency region of 500 to 4000 Hz.

Method of Evaluation

After completion of the construction of the Vowel Corrector, it was decided to run a number of tests in order to evaluate its proper functioning. It is obvious that a complete evaluation should consist of an assessment of the displaying characteristics of the equipment as well as its practicability as a teaching device for the deaf. In the following, we present the result of the former type of evaluation. Research with respect to the latter aspect is in progress.

The displaying characteristics of the Vowel Corrector, i.e. its segmentation and discrimination accuracy, were studied by observing the Corrector's behavior for a large number of /*I*/- and /*i*/ sounds, spoken by new male speakers in a variety of contexts. We hoped that the results of these tests would give us a basis for deciding whether speaker normalization is indicated for the use of the Vowel Corrector, and if so in what manner.

Material and Procedure

Each of twenty adult male speakers (normal hearing) pronounced 100 monosyllables which were recorded on tape. Half of the syllables

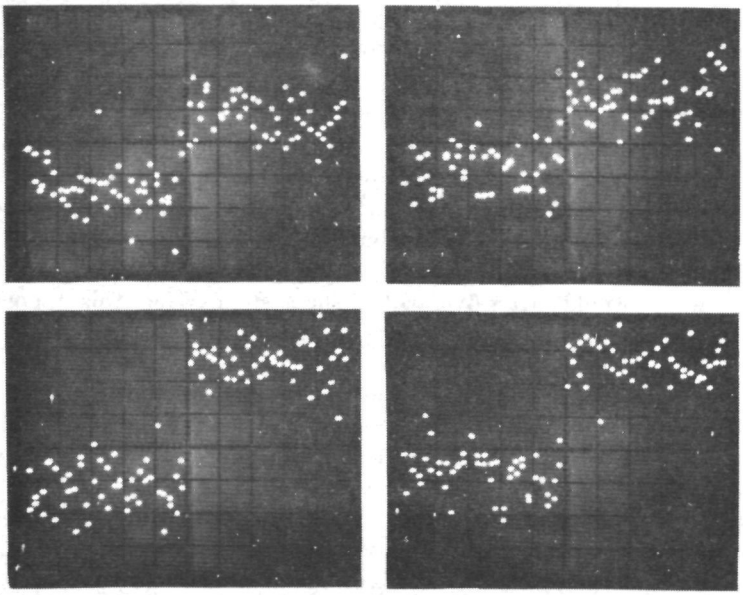

Fig. 11. Display of /*I*/ and /*i*/ sounds of 4 speakers (9-16-17-19) on the oscilloscope screen. The *Y*-axis displays the discriminant dimension. Left /*I*/, right /*i*/ vowsels

(CVCs) contained the vowel /*I*/, half the vowel /*i*/. The vowels appeared in 25 different consonantal contexts. All speakers sat at the same distance from the microphone and the setting of the input volume control was never changed during the recordings. The 2000 CVCs were fed one by one into the Vowel Corrector and displayed on the oscilloscope of which the *Y*-axis functions as discriminant dimension. In addition, the segmentation was made visible on a second oscilloscope (cf. Fig. 10); wrong segmentations were notified. For studying the displays, the following procedure was used. After the display of a vowel, the beam of the oscilloscope was shifted from left to right a little step. This was done in such a way that all 100 vowels of one speaker could be displayed on the screen at once: the /*I*/ vowels on the left-hand side, the /*i*/ vowels on the right-hand side. Moreover, each allophone had its own location on the screen. After having displayed all 100 vowels of one speaker the screen was photographed. This was repeated for all twenty speakers. Next, histograms were constructed, based on the photographs on which the *Y*-axis was divided into 20 equal intervals. Four photographs and four histograms based on them are shown in Figs. 11 and 12. The four speakers are representative for our sample.

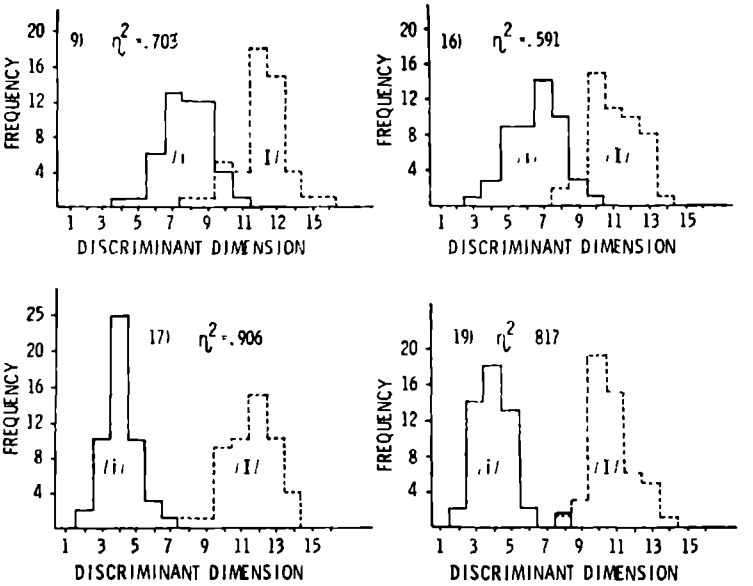

Fig. 12. Histograms based on the photographs of Fig. 10

Results

Segmentation

99.8% of the 2 000 segments is located in the central portion of the vowel, only 0.2% is located in the peripheral part of the vowel. These results indicate that our segmentation system works almost perfectly and should be feasible in all instances where vowel segmentation for speech training is desired.

Discrimination

A histogram of the displays of all 20 speakers together is shown in Fig. 13. The mean coordinate of the /*I*/ sounds in 12.07; the standard deviation is 2.33. The mean coordinate of the /*i*/ sounds is 6.14 with a standard deviation of 2.26. The total mean is 9.1. From the total variance 62.4% is explained by the differences between /*I*/ and /*i*/ and 12% by differences between speakers, the remaining 25.6% being interaction and error variance. First it was determined how well a discrimination could be obtained without any speaker normalization. Therefore a separation point on the discriminant dimension had to be chosen in such a way that the expected number of misprojections (i.e. an /*I*/ or /*i*/ projects on the wrong side of the separation point, or vice versa) is minimal. A second

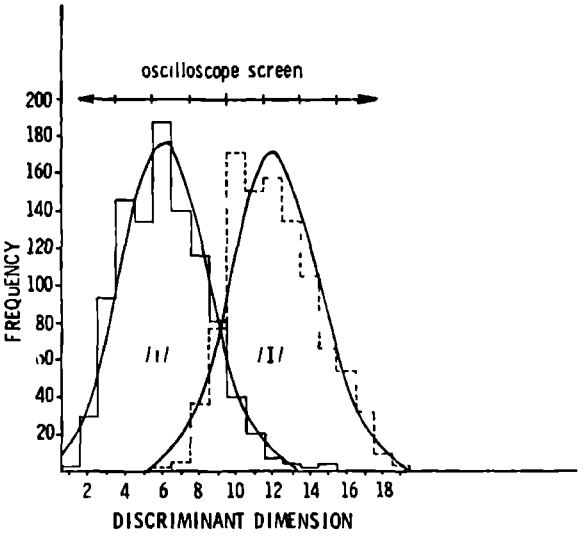

Fig 13 Histogram and normal distribution of all 1000 /*I*/ and 1000 /*ɪ*/ vowels as displayed on the oscilloscope screen

requirement for the selection of the separation point is that the increase in the number of errors either going to the right or to the left should be nearly equal This may be important because of possible fluctuations in the adjustment of the initial beam position on the *Y*-axis of the oscilloscope From the histogram in Fig. 12, however, it can be inferred that a relatively great number of /*I*/ sounds are located in the interval 9 5 to 10 5 In order to check whether the distributions of /*I*/ and /*ɪ*/ deviated much from the normal distribution, the cumulative proportion curves of both distributions were plotted on normal-probability paper Only very slight deviations from a straight line were found Therefore normal curves were constructed on both histograms (Fig 13) Since the variabilities of the two distributions differ only slightly ($\sigma_I = 2\ 33$, $\sigma_\iota = 2\ 26$), the cutting point of the two distributions is the most acceptable separation point The coordinate value of the cutting point on the discriminant dimension is 9 08 and the percentage of errors associated with this separation point is 9 7%. Thus one must conclude that the minimal percentage of errors with an a priori determined separation point, which means without any speaker-dependent correction, is 9 7% The most perfect speaker normalization is obtained when the separation point is determined a posteriori This means that the speaker first pronounces a large number of vowels and that then the optimal separation point is

Table 4 Optimal separation points on the discriminant dimension and corresponding number of errors per speaker

Speaker	Separation point	Number of errors in %	Speaker	Separation point	Number of errors in %
1	9.5	9	11	6.5	0
2	8.5	0	12	8.5	1
3	6.5	15	13	7.5	4
4	11.5	5	14	11.5	7
5	6.5	2	15	6.5	13
6	7.5	8	16	7.5	9
7	6.5	2	17	6.5	1
8	8.5	0	18	9.5	2
9	8.5	7	19	6.5	1
10	7.5	10	20	8.5	0

determined This was done for the 20 speakers and for each the corresponding number of errors was determined The result of this analysis are shown in Table 4 The separation points were always chosen on a boundary between two intervals There is a considerable spread of the individual separating points around the mean of 8 The mean percentage of errors is 4.8% Since this is a substantial improvement on 9.7%, a speaker dependent correction is quite desirable Of course, a perfect speaker normalization is not possible in practice, but a partial correction seems possible using the following procedure A speaker (deaf pupil), who is going to use the apparatus for training his pronunciation, is requested first to pronounce a number of CVCs with /I/ and /a/ On the basis of the resulting display it will be possible to determine broadly where the optimal separation point for this speaker will be Then the beam of the oscilloscope can be adjusted in such a way that this presumed separation point coincides with the horizontal midline of the oscilloscope screen This correction will certainly not be optimal, being based on a limited number of words Assuming that this adjustment will on the average not deviate more than a half division on the oscilloscope screen from the optimal correction, it is computed that such a deviation will, on the average correspond to an increase of errors from the optimal 4.8% to 7.5% This is still noticeably better than the 9.7% of errors associated with an a priori determined separation point Practice must show how well this method works

Discussion

It has been argued above that a number of special requirements are indicated for the construction of an articulation corrector It was asserted that it is not sufficient just only to display one or more aspects of the physical information, as is the case for the other types of visible speech

apparatus mentioned above, but that it is necessary to transform the physical information in such a way that the resulting display gives unambiquous and immediately understandable feedback to the subject. As a matter of fact an adequate articulation corrector should be a perfect automatic speech sound recognizer which discriminates speech sounds or aspects of speech sounds which are relevant for speech training. Because of the limited results in the field of automatic speech recognition it is not surprising that our Vowel Corrector has its limitations: it can only give feedback about a limited number of vowels at a time. It is, however, very well possible to make the apparatus suited for displaying other pairs or triads of vowels by changing the weighting coefficients. When the weightings for the vowels to be displayed are known, the only remaining thing to be done is to replace a set of resistors.

There is one point that still has to be mentioned. This concerns the supposed differences between our perceptual tolerance towards the pronunciation of certain vowels and the tolerance of the device. From our test for instance, it appears that a proportion of well-pronounced vowels is not as such identified by the device. In the same way it is possible that the Vowel Corrector "accepts" a vowel which is perceptually unacceptable. We have not yet examined this, but in the second part of the evaluation we will first of all turn our attention towards this aspect in a study in which the position of vowels pronounced by deaf boys will be correlated with the judged quality of these vowels. It may appear from this study that an improvement of the device is indicated. This might be the using of the second axis of the oscilloscope, now unused, for displaying a correct-incorrect dimension.

Next the practicibility of the device will be tested at the Institute for the Deaf at St. Michielsgestel. A first group of deaf boys who sometimes, but not always, pronounces both vowels correctly will try to improve their pronunciation by means of the Vowel Corrector. A second group of deaf boys who pronounce neither /*I*/ nor /*i*/ are going to try and acquire the pronunciation of these vowels via approximation, also using the information given by the Vowel Corrector. The subjects that will participate in the experiment will be older deaf boys with a fundamental voice frequency equalling that of adult male speakers. By using an appropriate research design we will try to isolate the effect of the Vowel Corrector so that the contribution of the device in the acquisition can be evaluated.

Many people have given very valuable assistance during this project. In particular I should like to thank Mr. L. C. W. Pols of the Institute for Perception at Soesterberg, to whom I owe many ideas for the realization of the Vowel Corrector and actual help during its development. Moreover I wish to acknowledge Professor E. Roskam for his advice on the computation and Mr. J. Drabbe for building the apparatus.

References

Annett, J.: Feedback and human behavior. Middlesex: Penguin Books 1969

Cattell, R. B. (Ed.) Handbook of multivariate experimental psychology. Chicago: Rand McNally & Company 1966

Cohen, M. L.: The ADL sustained phoneme analyzer. Amer. Ann. Deaf **113**, 247–252 (1968)

Ferber, L. A. Three parameter speech display. Conference record of the 1972 Conference on speech communication and processing, Boston, 1972. IEEE cat. no 72 CHO 596 7 AE, 252–257

Kalikow, D N, Swets, J. A.: Experiments with computer controlled displays in second-language learning. IEEE AU **20**, 23–28 (1972)

Kalikow, D N, Klatt, D. H.: Second language learning Report No 2008, Bolt, Beranek and Newman 1970

Kisner, J L, Weed, H. R.. The design of the visual vocoder Conference record of the 1972 Conference on speech communication and processing, Boston, 1972, IEEE cat. no 72 CHO 596-7, AE, 259–262

Klein, W., Plomp, R., Pols, L C W Vowel spectra, vowel spaces and vowel identification J Acoust. Soc. Amer 48, 999–1009 (1970)

McNeilage, P. F Motor control of serial ordering of speech. Psychol. Rev. **77**, 182 - 196 (1970)

Martony, J : Visual aids for speech correction. Summary of three years experience. Rep. Dept. of Speech Comm., Royal Inst. Techn., Stockholm 1969

Montgomery, G. W. G.: Communication factors related to the oscilloscope trace reading ability of profoundly deaf adolescents Paper 1970 Symposium on speech communication ability and profound deafness, Stockholm 1970

Nickerson, R. S., Stevens, L. N.. An experimental computer-based system of speech training aids for the deaf. Conference record of the 1972 Conference on speech communication and processing, Boston, 1972, IEEE cat. no. 72 CHO 596-7 AE 238–241

Nordmann, B. J.: A speech display simulation system. Conference record of the 1972 Conference on speech communication and processing, Boston, 1972, IEEE cat. no. 72 CHO 596-7, AE, 255–258

Pickett, J M., Constam, A.. A visual speech trainer with simplified indication of vowel spectrum. Amer. Ann. Deaf **113**, 253 - 258 (1968)

Plomp, R., Pols, L. C. W., Geer, J. P. van de Dimensional analysis of vowel spectra. J. Acoust. Soc. Amer. **41**, 707–712 (1967)

Pols, L. C. W., Tromp, H. R. C., Plomp, R. Frequency analysis of Dutch vowels from 50 male speakers J. Acoust. Soc Amer **53**, 1093–1102 (1973)

Potter, R K., Kopp, G. A, Green, H. C. Visible speech, New York Van Nostrand and Co 1947

Pronovost, W., Anderson, D, Lerner, R, Yenkin, L The development and evaluation of procedures for using the voice visualizer as an aid in teaching speech to the deaf Final Report proj no. 6-2017, Boston University, Aug 1967

Pronovost, W., Yenkin, L., Anderson, D C.: The voice visualizer. Amer. Ann. Deaf **113**, 230–239 (1968)

Risberg, A.: Visual aids for speech correction. Amer. Ann. Deaf **113**, 178–194 (1968)

Reich, S, Weed, H. R.: Evaluation of the visual vocoder in speech therapy. Conference record of the 1972 Conference on speech communication and processing, Boston, 1972, IEEE cat. no. 72 CHO 596-7 AE, 263–265

Sangster, F. L.: The "Bucket-Brigade Delay Line" a shift register for analog signals. Philips Technical Review **31**, 97–110

Schulte, K.: Optische Phonemdarstellung als Sprechgliederungshilfe fur horgeschadigte Kinder. Villingen: Neckar Verlag 1971

Scarson, M.: A speech training programme using the kamplex visible speech apparatus, Teacherof the Deaf **43**, 89—95 (1965)

Stark, R. E., Cullen, J. K., Chase, R. A.: Preliminary work with the new Bell telephone visible speech translator. Amer. Ann. Deaf **113**, 205—214 (1968)

Stark, R. E.: Teaching */ba/* and */pa/* to deaf children by means of real time spectral displays. Paper presented at the 79th Meeting of the Acoustical Society of America, Atlantic City 1970

D.-J. Povel
Psychologisch Laboratorium
Erasmuslaan 16
Nijmegen
The Netherlands

Psychol. Res. 37, 71 – 80 (1974)

Evaluation of the Vowel Corrector as a Speech Training Device for the Deaf

Dirk-Jan Povel
Department of Psychology, University of Nijmegen, The Netherlands

Received January 29, 1974

Summary. The Vowel Corrector, a device giving visual information about the identity of vowels spoken in monosyllabic words, was tested on a group of deaf boys. The description of the actual experiment is preceded by a number of considerations relating to the application of an articulation corrector in a speech therapy program for the deaf.

The deaf subjects participating in the experiment were divided into two groups: an experimental and a control group. The experimental subjects trained with the Vowel Corrector, while the subjects in the control group were trained by a professional speech trainer. At four moments in time data were collected with respect to the effect of training on pronunciation: before, during and directly after training, as well as 3 weeks later. On these occasions, recordings were made of the subjects pronouncing the list of words that was used during training, a list of words not used during training (transfer to other material) and a short text (transfer to another speech mode). Afterwards the recorded speech sounds were identified by two naive listeners. Subsequently the results were subjected to statistical analysis.

It could be concluded that the effect found in the experimental group is at least as large as the effect found in the control group which is considered a positive result. Implications of the findings are discussed.

The present article reports the use of the Vowel Corrector in the speech training of deaf pupils. The device is described in a preceding paper (Povel, 1974).

Introducing a visible speech apparatus into the speech learning situation of the deaf means adding a new dimension to it. In a sense, by doing this, the situation comes closer to that in which the normal hearing child acquires his speech. The normal hearing child for his speech acquisition essentially relies on a continuous comparison of speech in his environment with his own utterances. From these comparisons he infers the relations between the perceptual characteristics of speech sounds and their motor counterpart. This finally enables him to pronounce all elements of his native language in all occurring orders. This mode we shall call learning through output matching.

The way the deaf pupil is traditionally taught speech differs fundamentally from the process sketched above. Besides knowledge of results in terms of right-wrong statements, the main part of information given to

the deaf pupil consists in instructions concerning position and movements of his articulators. For his speech acquisition, therefore, the deaf pupil is primarily dependent on the indications that describe (partially) the process of speech production. This mode we shall call learning through process regulation.

Notice the difference: the hearing child hardly ever receives instructions concerning his articulators, whereas the deaf pupil never receives perceptual information about speech sounds (except the obviously very incomplete auditory information). In spite of the tremendous effort on the part of the teachers of the deaf, the speech of their pupils yet leaves much to be desired: it still sounds unnatural and it is difficult to understand for the naive listener. This must presumably be ascribed to the very high complexity of the speech process on the one hand and on the other to the impossibility of regulating this process through external instructions only. A visible speech apparatus restores to some degree the natural process: it gives the deaf subject the opportunity to see his own speech output, or more accurately a number of aspects of it, and the way this output varies with changes of the articulators.

Although some optimism is justified as regards the introduction of a visible speech apparatus into the speech learning situation of the deaf, one should still realize that some problems remain. In discussing these problems we will confine ourselves to the articulation corrector type of visible speech apparatus since the apparatus of which an evaluation is reported below, is a specimen of the latter type.

First of all one should bear in mind that the deaf pupil has never had the opportunity to develop his articulatory behavior in the way the hearing child has (we again neglect the contribution of the auditory information to the actual speech acquisition). Thus we do not know at all whether the deaf pupil will be able to employ the supplied information adequately.

Here a note should be made on the visual feedback that is given by an articulation corrector. Ideally an articulation corrector must give feedback that contains information with the help of which articulatory behavior can be controlled. The feedback however can only have a controlling function when the relations between the feedback and the motor domain are understood; and these relations have to be learned. Therefore during the first attempts that are based on trial and error the display fulfills only a reinforcing function by giving knowledge of results in terms of correct-incorrect. By comparing the changes in the visual display with changes in the motor commands to the speechorgans the subject may discover the relations between perceptual and motor correlates of speech sounds. From that time on the visual feedback can help him to guide his actions.

Secondly, there is the fact that speech acquisition is a long term process: it takes the hearing child several years to learn the relations between the acoustic and articulatory dimensions of speech and to apply them efficiently in his pronunciation. Several studies show that up till his eighth year of life, the hearing child keeps adding to his articulatory repertoire (Winitz, 1969). A consequence of this fact may be that the effectiveness of a speech corrector can only be established after a very long training period. And even then it will be almost impossible to supply the deaf subject with the same amount of information as that received by the hearing child.

Thirdly we should mention the limitations that are due either to the limited capacity of the eye in processing the visualized acoustic information or to the technical problems involved in the visual display of acoustic information. This type of limitations is first of all reflected by the fact that there are different visible speech correctors for the different aspects of the acoustic information that are traditionally distinguished: pitch and intonation, intensity, rhythm and articulation. The consequence of this fact, namely that the different aspects have to be trained independently will presumably not give rise to additional problems for the deaf since there is evidence that the hearing child also learns the different aspects more or less independently (Nakazima, 1962, Tonkova-Yampol'skaya, 1969).

The limitations of the latter type arise especially in the construction of articulation correctors. These devices have a rather complicated double function: on the one hand they must give a visual transformation of the acoustic information relevant to the correction or acquisition of articulation, on the other hand they must also indicate to what category (e.g. phonemes) the spoken sound belongs. In our articulation corrector the two functions could only be realized adequately if the number of sounds to be displayed at any one time were restricted to two (in the present study, /*I*/ and /*i*/). This measure which improves the discriminability (categorization power) of the apparatus, at the same time, however, reduces the view on the perceptual space: only the /*I*/ — /*i*/ dimension is shown. In this context we should mention a pilot study examining the similarity between the categorization of /*I*/ and /*i*/ sounds by the Vowel Corrector and by the human observer. In this study a number of monosyllables spoken by deaf boys were recorded on tape and subsequently both displayed on the Vowel Corrector and identified by two observers. Next the position on the display of each vowel was related to the perceptual identification. From this examination we could conclude that the similarity is very reasonable, be it that the tolerance of the apparatus is lower than perceptual tolerance. This means that a listener may accept a sound as /*I*/ or /*i*/, while the display is inconclusive (the spot appears somewhere near the midline). Sounds that are clearly displayed on the

Vowel Corrector as /*I*/ or /*i*/ are only very rarely perceptually unacceptable.

Finally we should mention a limitation that is directly connected to the Vowel Corrector. Since the display of this apparatus is based on the spectral characteristics of the incoming signal, it obviously can only aid in correcting those mistakes that are shown up in the spectral qualities of the sound. We shall return to this point at a later stage.

Method

Ten normal deaf boys from the "Instituut voor Doven" at St. Michielsgestel in the age group of 15 to 19 years with hearing losses varying from 98 to 125 dB ISO (average measures at 500, 1000 and 1500 Hz) were selected on the basis of pronunciation problems of /*I*/ and /*i*/ sounds in a variety of monosyllables and divided into two groups. The two groups were matched for age, audiogram, intelligence, initial level of errors in pronouncing /*I*/ and /*i*/ sounds and roughly on their general motivation to improve their pronunciation. The groups were named experimental (E) and control (C) group. Although we recognize the problem of forming two adequately matched groups of deaf boys we still preferred this design to a design in which the subjects are their own control and in which only the sounds to be learned differ in the experimental and control condition (Pronovost, 1967). The latter design has the disadvantage of severe potential a priori differences between the experimental and control condition which cannot be eliminated. The boys in the E group worked with the Vowel Corrector in a way which will be described presently, while the C group worked with a professional speech trainer. The speech trainer belonged to the permanent staff of speech therapists working at the institute where the research was conducted. Her teaching method may therefore be considered representative for the traditional speech training method as applied in the education of the deaf. We should mention that the speech trainer did not regard her task in the experiment as an extraordinary one. Indeed, in normal speech lessons exclusive prolonged attention will generally not be given to the production of merely two sounds. For the present study, however, this point is not relevant, since the experimental variable is the method of training used in the two groups. The two groups followed as far as possible the same plan during the ten 15 min training sessions. The E group had one extra session preceding the training in which the operation of the Vowel Corrector was explained and practised with the help of a very detailed instruction with many examples and exercises typed on cards.

The material used during the experiment consisted of three lists of words and a text. List I contained 38 CVCs arranged in 19 minimal pairs

with the vowel being either /*I*/ or /*i*/ e.g. /*bIt*/-/*bit*/. For the construction of the CVCs, 12 different initial consonants and 9 different final consonants were used. List II was made up of 32 monosyllabic meaningful words which were chosen on their estimated frequency in normal conversation. The words had either one, two or three initial consonants or final consonants, or both; half of the syllabic nuclei were formed by /*I*/, half by /*i*/. List III contained 34 CVCs, again half with vowel /*I*/ and half with vowel /*i*/. The three lists showed no overlap. The text, finally, consisted of a simple little story of approximately 100 words including 19 words with vowel /*I*/ and 23 with vowel /*i*/. Most of these words also occurred either in List I or List II.

During the training sessions the subjects of the E group were seated in front of the apparatus with a good view on the screen of the Vowel Corrector. They held a push button which could be pressed in order to show up a word on the display screen. The boys were wearing their usual hearing aids.

The operation of the Vowel Corrector and the display of speech signals on its screen are severely interfered with by sounds occurring just before or during the pronunciation of a word. Examples of such disturbing sounds are jogging the table on which the microphone is placed, scraping of feet on the floor, coughing, passing aeroplanes and thunder. These sounds either unduly trigger the microphone closing relais, or they make the display invalid. Therefore it was necessary to make the pupils aware of this fact and to work in a silent room.

Each session was divided into two periods. During the first period the subject used the minimal pairs of List I which were typed on cards and which were, during the earlier sessions, handed to him by the experimenter. The subject was instructed to try and pronounce the two words in such a manner that the spot on the screen corresponding to the /*I*/ sound would always appear below the spot corresponding to the /*i*/ sound. The larger the distance between the two spots the better. Note that it was only the relative position of the spots on the screen that mattered at this stage, not their absolute position. This procedure was chosen deliberately for the following two reasons. First, this task did not make excessively high demands on the part of the subject which was important in order to keep him motivated. Secondly, during this period, the experimenter had the opportunity to adjust the beam of the oscilloscope, if necessary, for speaker normalization. This could be done without any interruption of the progress of training. This adjustment was necessary only during the earlier sessions; at a later stage the optimal adjustment could be set before the training session started. In the second period of each session the subject used the words of List II. This time the task was to get the spot corresponding to either one of the two sounds on a

restricted region of the screen: /*I*/ below the midline and /*i*/ above the midline.

Now, typically, the subject made new attempts at a wrongly pronounced word (as indicated by the display) until its projection reached the desired position on the screen. If a subject had correctly pronounced a word (vowel) several times in succession, he was instructed to repeat the word several times without visual feedback (i.e. not to press the button) and to pay particular attention to the tactile and kinestetic sensation of his articulators. An occasional check was made by the subject by displaying the word on the Vowel Corrector's screen. The subjects received no further instructions during the training, indeed, from the outset they were encouraged to work on their own.

As mentioned earlier, the C group followed the same procedure as far as possible. Here again each session was divided into two periods: the first period being devoted to practising the pronunciation of the minimal pairs of List I and the second period to pronouncing correctly the words from List II. The only difference with the E group was that instead of receiving visual feedback from the Vowel Corrector, the subjects received articulatory instruction from a specialist in the speech education of the deaf.

Recordings of the pronunciation of all subjects were made at four points in time: 1. before the first training session, 2. after the fifth training session, 3. at the termination of training, 4. three weeks after termination of training. During these four tests the subjects read the words from the three lists and the text mentioned above. Thus for every subject we collected on each occasion 106 monosyllabic words spoken in isolation, half with vowel /*I*/, half with vowel /*i*/, plus 42 /*I*/ and /*i*/ sounds spoken while reading a text. From these recordings, four measures concerning the effect of training could be inferred: 1. improvement in the pronunciation of the words used during training (List I and II), 2. transfer of learning to other words (List III), 3. transfer to another mode of speech (Text), 4. retention of the effect (test three weeks after termination).

For purposes of scoring all vowels had to be judged for intelligibility. Identification, rather than quality, was considered to be the most relevant measure for the correctness of pronunciation. Therefore two observers, unfamiliar with deaf speech, were instructed to identify the words that were presented to them in the judging sessions and to write these down. Identifying whole words is closer to the normal way of perceiving speech than identifying only the vowel section of words. The relevant speech sounds in the text were judged by the observers while listening to the story as it was spoken by the deaf pupils. Thus the observers could make

use of the constraints of the story while identifying the /*I*/ and /*i*/ sounds. This procedure was chosen deliberately since it represents again a better approximation of the natural way of perceiving speech. Later, the responses of the judges were compared with the corresponding words presented to the deaf subjects during the test. Only when both judges had correctly identified the intended vowel, was the pronunciation categorized as correct. The average percentage of interjudge agreement was 89.

Results

The Vowel Corrector in Operation

Before going into details of the data analysis some comments should be made about the use of the Vowel Corrector during training. The subjects of the E group appeared to have understood the instructions: they had no trouble operating the apparatus and showed good comprehension of the meaning of the display. The operation of the apparatus itself gave no trouble neither did segmentation except in the case of one subject who occasionally spoke an initial /*r*/ (frontal tongue *r*) louder than the subsequent vowel. For this subject the words with initial /*r*/ were replaced.

Next a comment should be made on the production of vowels not being /*I*/ or /*i*/ and their display on the Vowel Corrector. From an earlier investigation of speech errors of deaf boys we learnt that the errors in the production of /*I*/ and /*i*/ sounds in monosyllables can be divided into two categories. The first and largest category comprises reversals from /*I*/ and /*i*/ and the production of speech sounds perceptually intermediate between /*I*/ and /*i*/. The second category consists of quite different vowels e.g. /*ɛ, e, u, y, o*/. The latter category constituted approximately 6% of all the errors: we also found this value in our E and C group results. This latter type of vowels, except for /*y*/, projected outside the screen, which meant that no spot was displayed on it. It was explained to the subjects that when this happened the vowel produced was rather deviant and they were instructed to make another attempt. The vowel /*y*/, which was produced rarely, was generally projected near the midline and thus meant for the subject a sound that had to be corrected. This matter did not give rise to any further problems during training.

Data Analysis

Fig. 1 shows the improvement over time of the E and the C group for the different materials in three separate diagrams.

The first diagram (A) shows the learning curves of both groups for the material that was used during training; the second pair of curves (B)

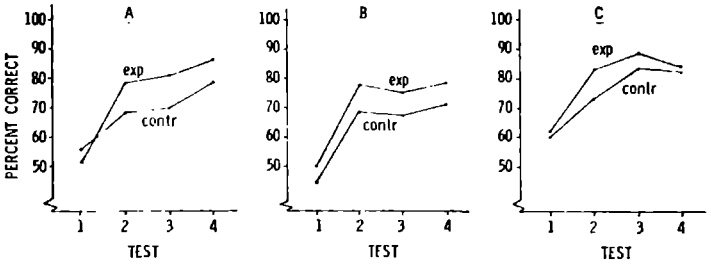

Fig. 1. Improvement over time of E and C group. A. Words used in training; B. Words not used in training; C. Text

show the transfer to other single words, while the last diagram (C) shows transfer to words spoken while reading text. The most important impression from these diagrams is that both groups appear to improve their pronunciation. Next there appears a tendency for the E group to perform better than the C group. An other noteworthy characteristic is that the effect of the training of both the E and the C group appears to have reached its final level by the end of the fifth training session.

Transfer to not learned words, as computed from the posttraining test, amounts to 91% and 94% of the effect on the learned words for the E and the C group respectively.

Because of the differences in judgement procedures used for the words spoken in isolation and for the words spoken in the text these two categories were analyzed separately. An analysis of variance performed on the data of the isolated words reveals that of the three main effects only Moment in time is significant ($F3,21 = 18.9$; $p < 0.001$). For the difference between E and C group and between the two types of material non significant F values are found. None of the interactions appears to be significant.

An analysis of variance performed on the data of the words spoken in text yields similar results: here too only Moment in time is found to reach significance ($F\,3,24 = 37.3$; $p < 0.001$).

Discussion

In view of the reported results we have reason to be optimistic with respect to the application of the Vowel Corrector. We may conclude that it at least matches a speech trainer as far as the correction of / *I* / and / *i* / sounds is concerned. Transfer of training both to other words and to another mode of speaking appear to occur to a satisfactory degree. The effect of training is retained over a period of at least 3 weeks.

More importantly, the results indicate the possibilities of another method of teaching speech to the deaf, which, in the introduction we have termed learning through output matching In this method the articulatory behavior of the pupil is not shaped with the help of verbal instructions concerning position and movement of the articulators as is the case in the traditional teaching method, but instead the articulatory behavior is indirectly guided by visual feedback This feedback, being a transformation of some important information bearing aspects of the acoustic speech signal apparently can to some extent take over the assumed function of the ear in the speech acquisition process of the hearing child This function might best be described as providing a norm plus an indication of the degree to which the subject's speech product matches this norm In order for this information to be useful, the subject is assumed to relate the visual display to articulatory dimensions and to form transformation rules which enable him to use the visual information for future corrections From our study we might tentatively conclude that the subjects do appear to be able to use the information in the described manner This conclusion plus the fact that withdrawal of the feedback is not directly followed by a deterioration of performance should be considered the most important results of this research, showing real perspectives for this alternative speech training method It supports our view that articulatory acquisition can and should be based on a perceptual, rather than on a motor criterion

The research also shows that further developments are needed First the Vowel Corrector should be made to display also other pairs of vowels this can be accomplished rather easily (see the preceding paper) Secondly in a next phase the Vowel Corrector must be developed towards a general articulation corrector, i e one that can also display consonants This will, to a large extent, be feasible in view of the recent results in automatic phoneme recognition, but problems will certainly arise with the development of an adequate segmentation procedure The third modification will presumably appear to be the hardest to realize It concerns an extension of the information given by the apparatus so that it becomes applicable to the correction of other classes of articulation errors than studied thus far With respect to this point we should mention a type of articulation error that is not signalled adequately on the Vowel Corrector pronouncing a diphthong instead of a vowel When a diphthong containing the intended vowel is pronounced, it will depend on the part of the diphthong that happens to be sampled, whether the display indicates a deviation from the norm

Certainly the future developments of the Vowel Corrector to a full-grown articulation corrector will meet with several problems We are convinced however that it is worth the effort to solve these problems

since we believe that devices displaying aspects of the speech product visually have promising perspectives for the speech training of the deaf.

The author wishes to thank Mrs. D. Vughts-Sterke for the conscientious way in which she ran the control group and Eddy Buermans, Karel Ettinger, Bert Heymans, Peter Jeurnnck, Dré op den Kamp, Emiel Klaassen, Juup Mollee, Henny Queens, Hans Roeven, Bertie Staal all from the Instituut voor Doven at St. Michielsgestel, for their kind cooperation in the experiment. The author is also indebted to Prof Levelt for reviewing an earlier draft of this paper.

References

Nakazima, S.: A comparative study of the speech developments of Japanese and American English in childhood. Stud. Phonol. **2**, 27—39 (1962)

Povel, D. J.: Development of a vowel corrector for the deaf. Psychol. Res. **37**, 51 - 70 (1974)

Pronovost, W., Anderson, D., Lerner, R., Yenkin, L.: The development and evaluation of procedures for using the voice visualizer as an aid in teaching speech to the deaf. Final Report proj. no. 6-2017. Boston University, aug. 1967

Tonkova Yampol'skaya, R. V.: Development of speech intonation in infants during the first two years of life. Sov. Psychol. Psychat. **7**, 48 54 (1969)

Winitz, H.: Articulatory acquisition and behavior. New York: Appleton Century Crofts 1969

D.-J. Povel
Psychologisch Laboratorium
Erasmuslaan 16
Nijmegen
The Netherlands

SUMMARY

Any study intending to evaluate the various applications of 'visible speech apparatus' for speech training of the deaf (viz. articulation correctors) must first of all attempt to reveal the relevant aspects of the process of speech production. Secondly, it must assess the differences between the process of speech acquisition of the hearing on the one hand and of the deaf on the other. Thirdly, it must try to assign a proper place to the rather deviant form of presenting information by means of a visible speech apparatus.

1. In the present study, Chapter 1 is devoted to the process of speech production. For the purpose of our discussion we have looked for an analogy in order to make the complexities of speech production more explicit. As a useful analogy we propose the mechanisms that underlie flute playing. During the elaboration of the analogy, we come across a number of aspects in flute-playing behaviour that supposedly have their counterpart in the process of speaking. Thus we look into (i) the issue of the unit of speech production; (ii) the problem of the transformations occurring between input and output; (iii) the question concerning the role of feedback. The main part of the chapter is, therefore, devoted to a detailed discussion of the literature covering these aspects of speech production. In this context, two alternative views on speech production are critically reviewed.

Of great significance to our purpose is that the flute-playing model of speech production appears to have sufficient descriptive validity. Most importantly, it stresses the fact that the one-to-one relationship between perceptual units and neural commands – as it has been proposed in various forms in the literature – is inadequate. Indeed, there is growing evidence which indicates that one speech sound can be produced by means of several different articulatory patterns which, in turn, are caused by, equally different, neural commands. Moreover, the model suggests that, depending on the preceeding speech sound, different articulatory behaviour is required.

These findings have a bearing not only on the assessment of articulatory performance but also – and this is of greater practical import-

ance – on speech training.

2. In Chapter 2, two different modes of speech acquisition are compared. These have been termed 'learning through output matching' and 'learning through process shaping'. The former represents the manner in which the hearing child acquires speech; the latter describes the way the deaf child traditionally learns to speak.

A number of essential shortcomings of the latter method are mentioned and a plea is made for the alternative method in which information about the speech product is given. Thus, the subject is provided with an appropriate norm. Moreover, the deaf subject is given the opportunity to make use of the natural matching procedures which form the kernel of speech acquisition by the hearing.

3. Chapter 3 offers a detailed description of the development of the Vowel Corrector and its final design. The Vowel Corrector displays vowels spoken in monosyllables as light spots on an oscilloscope screen. Because of the limited number of dimensions available, only a restricted number of vowels can be displayed in a single session without losing discriminatory power. The locus of the spot on the screen forms the criterion for a correct pronunciation.

4. In the fourth chapter, finally, the functioning of the Vowel Corrector as a speech training aid for the deaf is evaluated. For this purpose, two matched groups of deaf boys received speech training. One group was trained by means of the Vowel Corrector, the other received instruction from a professional speech trainer. The improvement in intelligibility of the vowels practised was assessed for both groups. It could be concluded that the effect due to the Vowel Corrector is at least as large as the effect due to the speech trainer, which is considered an encouraging result.

SAMENVATTING

Het onderzoek beschreven in dit proefschrift behelst een theoretische en praktische waardering van de mogelijkheden van visueel afgebeelde acoustische informatie omtrent het spraaksignaal ten behoeve van het spraakonderwijs aan doven. Het eerste en tweede hoofdstuk zijn gewijd aan het theoretische facet, het derde en het vierde aan het praktische.

1. Het eerste hoofdstuk gaat in op de vraag wat spreken, het produceren van spraak, eigenlijk is. Teneinde een duidelijker beeld te ontwikkelen van het komplexe mechanisme dat zorg draagt voor een juist besturing van de artikulatoren gedurende het spreken is een analogie uitgewerkt die in een aantal opzichten gelijkenis vertoont met dit gedrag het bespelen van een fluit. Deze analogie brengt ons op het spoor van een aantal deelprocessen die vermoedelijk ook een rol spelen bij het produceren van spraak. Voor de beantwoording van de vraag in hoeverre de voorgestelde analogie een acceptabel model oplevert, worden onder meer de volgende punten aan de orde gesteld. (i) Welke is de 'eenheid' van spraakproduktie? (ii) Welke processen voltrekken zich tussen invoer en uitvoer binnen het spraakproduktiemechanisme? In dit verband worden twee alternatieve modellen kritisch besproken. (iii) Wat is de rol van feedback bij het spreken? Dit laatste punt is onder meer van belang in verband met de vraag in hoeverre de dove in staat geacht mag worden zijn spraakvermogen te ontwikkelen. Het grootste deel van het hoofdstuk is gewijd aan de behandeling van de literatuur die betrekking heeft op deze aspekten van het spreken. Hieruit blijkt dat het voorgestelde model de gereleveerde data adekwaat kan beschrijven. Het model suggereert met name dat de opvatting van een één-op-één-relatie die zou bestaan tussen fonemen enerzijds en neurale kommando's voor de artikulatoren anderzijds – zoals in verschillende varianten in de literatuur voorgesteld – onjuist is. Inderdaad wordt er steeds meer evidentie aangevoerd die erop wijst dat één spraakklank met behulp van verschillende artikulatoire konfiguraties kan worden gerealiseerd, waarbij geldt dat deze verschillende konfiguraties zijn voortgebracht door verschillende neurale kommando's en niet kunnen worden toegeschreven aan perifere effekten Het voorgestelde model laat tevens zien dat, afhankelijk van de voorafgaande spraakklank, verschillende aktiviteit nodig is voor de realisatie van een omschreven artikulatoire konfiguratie.

Het model heeft niet alleen gevolgen voor de wijze waarop artikulatieontwikkeling moet worden bepaald, tevens volgen er aanwijzingen uit voor de praktijk van het spraakonderwijs.

2. In hoofdstuk 2 worden twee verschillende methoden van spraakverwerving met elkaar vergeleken. De eerste beschrijft de wijze waarop de horende leert spreken, de tweede is de methode die in het spraakonderricht aan doven traditioneel gevolgd wordt. Een aantal essentiele tekortkomingen en problemen samenhangende met deze laatste methode wordt genoemd. Als oplossing voor deze problemen wordt gepleit voor een alternatieve weg waarbij de dove informatie over het spraakprodukt zelf wordt gegeven. Op deze wijze wordt de dove voorzien van een adekwate norm welke hem in staat stelt gebruik te maken van de natuurlijke gelijkmakingsprocedures die geacht worden de kern van de normale spraakontwikkeling te vormen.

3. Hoofdstuk 3 geeft, uitgaande van een serie eisen die aan een artikulatiekorrektor moeten worden gesteld, een gedetailleerde beschrijving van de ontwikkeling van de klinkerkorrektor. De klinkerkorrektor beeldt klinkers, gesproken in monosyllaben, af als lichtpunten op het scherm van een oscilloscoop. De plaats op het scherm vormt het kriterium voor een juiste uitspraak. Vanwege het beperkt aantal dimensies dat ter beschikking staat kunnen er tegelijkertijd niet meer dan twee tot drie klinkers afgebeeld worden zonder verlies van diskriminatie. Het apparaat kan echter gemakkelijk ingesteld worden voor het trainen van andere klinkers.

4. In het vierde hoofdstuk wordt de klinkerkorrektor geevalueerd als een hulpmiddel voor de spraaktraining van doven. Voor dit doel ontvingen twee vergelijkbare groepen van dove jongens spraaktraining. De ene groep trainde met de klinkerkorrektor terwijl de andere groep aanwijzingen ontving van een logopedist verbonden aan een doveninstituut. De verbetering van de verstaanbaarheid van de geoefende klinkers werd bepaald voor beide groepen. Hieruit kon worden gekonkludeerd dat het positieve effekt dat werd geobserveerd in de groep, die trainde met de klinkerkorrektor, minstens zo groot is als het effekt dat werd bereikt in de andere groep. Dit resultaat mag worden beschouwd als veelbelovend ten aanzien van de verdere ontwikkeling van apparatuur voor de visuele afbeelding van spraak en de toepassing ervan in het onderwijs aan doven.

REFERENCES

(This list only contains the references of Chapters 1 and 2)

ABBS J.H.: The influence of the gamma motor system on jaw movements during speech: a theoretical framework and some preliminary observations. *J. Speech Hear. Res. 16*, 175-200 (1973).

ALLPORT F.H.: *Social Psychology*. Boston, Houghton. 1924.

BELL V.L.: *Sensorimotor learning. From Research to Teaching*. Pacific Palisades Cal., Goodyear Publ. Comp. Inc. 1970.

BERRY M.F.: *Language disorders of children: the bases and diagnoses*. New York, Appleton. 1969.

BISHOP M.E., RINGEL R.L., HOUSE A.S.: Orosensory perception, speech production and deafness. *J. Speech Hear. Res. 16*, 257-266 (1973).

BOUHUYS A. (Ed.): Sound production in man. *Annals New York, Acad. of Sciences. 155*, art. 1. 1968.

BRAINE M.D.S.: The acquisition of language in infant and child. In: Reed C.E.: *The learning of language*. New York Appleton Cent. Croft 1971.

CARROL J.B.: *Language acquisition, bilingualism and language change*. Encycl. Educ. Res. New York, MacMillan 1960.

CHASE R.A.: Abnormalities in motor control secondary to congenital sensory deficits. In: Bosma J.F. (Ed.): *Symposion on oral sensation and perception*. Illinois, Springfield. 1967.

COHEN A.: Versprekingen als verklappers van het proces van spreken en verstaan. *Forum der Letteren 6*, (1965).

CRAIK K.J.W.: Theory of the human operator in control systems. I. The operator as an engineering system. *Brit. J. Psych. 38*, 56-61 (1947).

EISENSON J., AUER J.J., IRWIN J.V.: *The psychology of communication*. New York, Appleton. 1963.

FAIRBANKS G.: A theory of the speech mechanism as a servosystem. *J. Speech Hear. Dis. 19*, 133-139 (1954).

FAIRBANKS G.: Selective vocal effects of delayed auditory feedback. *J. Speech Hear. Dis. 20*, 333-346 (1955).

FITTS P.M., POSNER M.I.: *Human Performance*. Belmont Cal. Brooks and Cole. 1967.

FROMKIN V.A., LADEFOGED P.: Electromyography in speech research. *Phonetica 15*, 219-242 (1966).

FROMKIN V.A.: Neuromuscular specification of linguistic units. *Lang. Speech 9*, 170-199 (1966).

FROMKIN V.A.: The non-anomalous nature of anomalous utterances. *Language 47*, 27-52 (1971).

FRY D.B.: The development of the phonological system in the normal and the deaf child. In: Smith F, Miller G.A. (Eds): *The genesis of language*. Cambridge Mass. M.I.T. Press 1966.

GIBBS C.B.: Probability learning in step-input tracking. *Brit. J. Psych. 56*, 233-242 (1965).

GOLDMAN-EISLER F.: On the variability of the speed of talking and on its relation to the length of utterances in conversation. *Brit. J. Psych. 45*, 94-107 (1954).

GRANIT R.: *The basis of motor control*. London, Acad. Press 1970.

GREGOIRE A.: *L'apprentisage du language*. l'Université de Liège. Liège, 1937.

HEFFNER R.M.S.: *General Phonetics*. Madison, University of Wisconsin Press 1969.

HORII Y., HOUSE A.S., KUNG-PU LI, RINGEL R.: Acoustic characteristics of speech produced without oral sensation. *J. Speech Hear. Res. 16*, 67-77 (1973).

INGRAM D.: Phonological rules in children. *J. Child Lang. 1*, 49-64 (1974).

IRWIN O.C.: Research on speech sounds for the first six months of life. *Psych. Bull. 38*, 277-285 (1941).

IRWIN O.C., CHEN H.P.: Development of speech during infancy: curve of phonemic types. *J. Exp. Psych. 36*, 431-436 (1946).

JAKOBSON R.: *Child Language, aphasia and phonological universals*. Den Haag, Mouton 1968.

KALIKOW D.N., KLATT D.H.: Second language learning. *Report nr. 2008. Bolt Beranek and Newman*. 1970.

KOOPMANS VAN BEINUM F.J.: Formantfrequenties en duur van klinkers in

in verbonden spraak. In: *Verslagen van de voorjaarsvergadering van de nederlandse vereniging voor fonetische wetenschappen.* 1972.

KREMERS A.: De akwisitie van spraakklanken in prelinguale en linguale fase. Nijmegen, Unpubl. 1972.

LADEFOGED P.: *Three areas of experimental phonetics.* London, Oxford Press 1967.

LASHLEY K.S.: The problem of serial order in behavior. In: Jeffres L.A. (Ed.): *Cerebral mechanisms in behavior.* New York Wiley 1951.

LEE B.S.: Effects of delayed speech feedback. *J. Acoust. Soc. Amer. 22,* 824-826 (1950).

LEE B.S.: Some effects of side-tone delay. *J. Acoust. Soc. Amer. 22,* 639-640 (1950).

LEE B.S.: Artificial stutter. *J. Speech Hear. Dis. 16,* 53-55 (1951).

LENNEBERG E.H.: Understanding language without ability to speak: a case report. *J. Abnorm. Soc. Psych. 65,* 419-425 (1962).

LENNEBERG E.H.: Speech as a motor skill with special reference to non-aphasic disorders. In: *Monographs of the society for research in child development. 29,* 115-127 (1964).

LENNEBERG E.H.: *Biological foundations of language.* New York, Wiley 1967.

LEVELT W.J.M.: *What has become of LAD?* Den Haag, Peter de Ridder Press 1975 i.p.

LEWIS M.M.: *Language thought and personality in infancy and childhood.* London, Harrap 1963.

LIBERMAN A.M.: Some results of research on speech perception. *J. Acoust. Soc. Amer. 29,* 117-123 (1957).

LIBERMAN A.M., COOPER F.S., HARRIS K.S., MCNEILAGE P.F.: A motor theory of speech perception. *Proceed. of the speech comm. seminar.* Stockholm, Royal Inst. Techn. 1963.

LIBERMAN A.M., COOPER F.S., SHANKWEILER D.P., STUDDERT-KENNEDY M.: Perception of the speech code. *Psych. Rev. 74,* 431-461 (1967).

LINDBLOM B.E.F.: Spectrographic study of vowel reduction. *J. Acoust. Soc. Amer. 35,* 1773-1781 (1963).

LINDBLOM B.E.F., Sundberg J.: Neurophysiological representation of speech sounds. *Paper at the XVth World Congress of logopedics and phoniatrics.* Buenos Aires, Argentina 1971.

LINSENER J., LINSENER H.J.: Untersuchungen zum Lee-effekt I. *Zeitschrift f. Psych. 168*, 26-58 (1963).

LOCKE J.L.: Oral perception and articulation learning. *Percept. Mot. Skills 26*, 1259-1264 (1968).

LOCKE J.L.: Ease of articulation. *J. Speech Hear, Res. 15*, 194-200 (1972).

MacKAY D.G.. Spoonerisms: the structure of errors in the serial order of speech. *Neuropsychologia 8*, 323-350 (1970).

MacNEILAGE P.F., DeClerk J.L.: On the motor control of coarticulation in CVC monosyllables. *J. Acoust. Soc. Amer. 45*, 1217-1233 (1969).

MacNEILAGE P.F.: Motor control of serial ordering of speech. *Psychol. Rev. 77*, 182-196 (1970).

McCARTHEY D.: Language development in children. In: Carmichael L.: *A manual of child psychology*. New York, Wiley 1966.

McCLEAN M.: Forward coarticulation of velar movement at marked juncture boundaries. *J. Speech Hear. Res. 16*, 286-296 (1973).

McNEIL D.: *The acquisition of language. The study of developmental psycholinguistics*. New York, Harper & Row 1970.

MENYUK P.: *The acquisition and development of language*. London, Prentice-Hall. 1971.

MILLER G.A., NICELY P.E.: An analysis of perceptual confusions among some english consonants. *J. Acoust. Soc. Amer. 27*, 338-352 (1955).

MILLER G.A., GALANTER E., PRIBRAM K.H.: *Plans and the structure of behavior*. New York, Holt 1960.

MORTIMER E.M., AKERT K.: Cortical control and representation of fusimotor neurons. *Am. J. Phys. Med, 40*, 228-248 (1961).

MYSAK E.D.: *Speech pahtology and feedback theory*. Springfield Illinois, Thomas 1966.

NOOTEBOOM S.G.: The tongue slips into patterns. *Nomen: Leyden studies in linguistics and phonetics*. 114-132, 1969.

NOOTEBOOM S.G.: The target theory of speech production. *I.P.O. Ann. Progr. Rep. 5*, 51-55 (1970).

NOOTEBOOM S.G., SLIS I.H.: A note on the degree of opening and the duration of vowels in normal and 'pipe' speech. *I.P.O. Ann. Progr. Rep. 5*, 55-58 (1970).

ÖHMAN S.E.G.: Coarticulation in VCV utterances. Spectrographic measurements. *J. Acoust. Soc. Amer. 39*, 151 - 168 (1966).

OLMSTED D.L.: A theory of the child's learning of phonology. *Language 42*, 531-535 (1966).

OLMSTED D.L.: *Out of the mouth of babes.* Den Haag, Mouton 1971.

POTTER R.K., KOPP G.A., GREEN H.C.: *Visible speech.* New York, van Nostrand 1947.

PRIBRAM K.H.: *Languages of the brain.* New York, Prentice-Hall 1971.

RINGEL R.L., HOUSE A.S., BURK K.W., DOLINSKY J.P., SCOTT C.M.: Some relations between oral sensory discrimination and articulatory aspects of speech production. *J. Speech Hear. Dis. 35*, 3-11 (1970).

RINGEL R.L., STEER M.D.: Some effects of tacilte and auditory alterations on speech output. *J. Speech Hear. Res. 6*, 369-378 (1970).

RUTHERFORD D.: Auditory-motor learning and the acquisition of speech. *Am. J. Phys. Med. 64*, 245-251 (1967).

SCHAERLAEKENS A.M.: *The two-word sentence in child language development.* Den Haag, Mouton 1973.

SCOTT C.M., RINGEL R.L.: Articulation without oral sensory control. *J. Speech Hear. Res. 14*, 804-818 (1971).

SIMON J.R.: Effect of ear stimulated on reaction time and movement time. *J. Exp. Psychol. 78*, 344-346 (1968).

SKINNER B.F.: Verbal Behavior. New York, Appleton Cent. Croft 1957.

SMITH F., MILLER G.A. (Eds.): *The genesis of language.* A Psycholinguistic approach. Proceedings of a conference on 'language development in children'. Cambridge Mass. M.I.T. Press. 1966.

SMITH K.U.: *Delayed sensory feedback and behavior.* London, Saunders Comp. 1962.

SMITH N.V.: *The acquisition of phonology: a case study.* London, C.U.P. 1973.

SNOW K.A.: A detailed analysis of articulation responses of 'normal' first grade children. *J. Speech Hear. Res. 6*, 277-290 (1963).

SNOW K.A.: A comparative study of sound substitution used by 'normal' first grade children. *Speech Monogr. 31*, 135-141 (1964).

STAATS A.W., STAATS C.K.: A comparison of the development of speech and reading behavior with implications for research. *Child Develop.* 33,

831-846 (1962).

STAATS A.W., STAATS C.K.: *Complex human behavior.* A systematic extension of learning principles. New York, Holt 1963.

STETSON R.H.: *Motor phonetics. A study of speech movements in action.* Amsterdam, North Holland 1951.

STRENGER F.: Radiographic, palatographic and labiographic methods in phonetics. In: Malmberg B. (Ed.): *Manual of phonetics.* Amsterdam, North Holland 1968.

SUSSMAN H.M.: What the tongue tells the brain. *Psychol. Bull. 77*, 262-272 (1972).

SUSSMAN H.M., MCNEILAGE P.F., HANSON R.J.: Labial and mandibular dynamics during the production of bilabial consonants: preliminary observations. *J. Speech Hear. Res. 16*, 397-420 (1973).

TEMPLIN M.C.: The study of articulation and language development during the early school years. In: Smith F. and Miller G.A. (Eds.): *The genesis of language.* Cambridge Mass. M.I.T. Press 1966.

UDEN A.M.J. van: *A world of language for deaf children.* St. Michielsgestel. Uitgave van het Instituut voor Doven 1968.

UDEN A.M.J. van: *Dove kinderen leren spreken.* Rotterdam, Universitaire Pers 1974.

WELLMAN B.L., CASE I.M., MENGERT I.G., BRADBURY D.E.: Speech sounds of young children. *University of Iowa studies in Child Welfare 5*, 2 (1931).

WICKELGREN W.A.: Context sensitive coding, associative memory and serial order in (speech)behavior. *Psychol. Rev. 76*, 1-15 (1969).

WINITZ H., PREISLER L.: Discrimination pretraining and sound learning. *Percept. Mot. Skills 20*, 905-916 (1965).

WINITZ H.: *Articulatory acquisition and behavior.* New York, Appleton Cent. Croft 1969.

WOODWORTH R.S.: The accuracy of voluntary movement. *Psychol. Rev. Monograph Suppl. 3*, 2 (1899).

Curriculum vitae

Dirk Jan Povel werd geboren op 29 oktober 1940 te Breda. Hij studeerde psychologie aan de katholieke universiteit te Nijmegen van 1961 tot 1967. Na het kandidaatsexamen koos hij de hoofdrichting funktieleer. Zijn afstudeeronderwerp had betrekking op de perceptuele interäktie van tegelijkertijd aangeboden optische en acoustische stimuli. Na zijn afstuderen trad hij halftime in dienst bij de afdeling kinderaudiologie van het Radboudziekenhuis te Nijmegen, en halftime bij het psychologisch laboratorium van dezelfde universiteit. Na een jaar ging hij full-time werken bij het psychologisch laboratorium. In 1970 werden kontakten gelegd met Dr. A. van Uden van het Instituut voor Doven te St. Michielsgestel, waaruit het onderzoek beschreven in deze dissertatie is voortgekomen.

STELLINGEN

1.

In het spraakonderricht moet de methode die tracht de spraakontwikkeling te bevorderen middels het verstrekken van informatie over het proces van het spreken onderscheiden worden van de methode die ditzelfde doel tracht te bereiken door informatie over het spraakprodukt te verschaffen.

2.

De verdere ontwikkeling van 'artikulatiekorrektoren' ten behoeve van spraakkorrektie van doven moet zich niet richten op het ontwerpen van apparatuur die standen of bewegingen van de artikulatoren zichtbaar maakt, maar op de konstruktie van apparaten die artikulatieaspekten afbeelden die opgesloten liggen in het akoestisch spraaksignaal.

3.

De moeilijkheid om een vreemde spraakklank korrekt uit te spreken wordt niet zozeer veroorzaakt door een motorisch onvermogen om de benodigde artikulatoire positie te realiseren dan wel door een perceptueel onvermogen om de relevante aspekten van de klank te identificeren.

4.

Spraak van doven wordt niet beter verstaanbaar wanneer duur en duurverhoudingen van de spraaksegmenten langs kunstmatige weg worden genormaliseerd.

5.

Het experimentele paradigma waarin de akoestische feedback de spreker vertraagd bereikt, is geen vruchtbare methode voor de bestudering van de rol van akoestische feedback in het proces van het spreken.

Het fenomeen dat een spraaksignaal waarvan alleen de energierijke gedeelten (klinkers) waarneembaar zijn doordat de zwakkere gedeelten zijn onderdrukt, aanzienlijk verstaanbaarder wordt door er ruis aan toe te voegen, kan slechts adekwaat worden beschreven in een spraakperceptiemodel dat een analyse-door-synthese-principe insluit.

Cherry C., Wiley R.I. *Nature, 214*, 1164 (1967)
Holloway C.M. . *Nature, 226*, 178 (1970)

7.

Het onderzoek dat middels het uitschakelen of vervormen van afzonderlijke parameters in het spraaksignaal het relatieve belang van deze afzonderlijke aspekten voor de normale spraakperceptie wil bepalen, houdt onvoldoende rekening met het adaptief vermogen van het waarnemingssysteem.

8.

De volgens de wet van Fechner te verwachten lineaire relatie tussen de snelheid waarmee een stuk muziek wordt uitgevoerd en de grootte van de geproduceerde verschillen in toonduur wordt slechts bij sommige musici aangetroffen.

9.

Het gebruik van het notenschrift in het muziekonderricht en tijdens het uitvoeren van muziek dient tot het minimum beperkt te worden waardoor een sturing van het muzikale gedrag door een auditieve representatie van de muziek in de plaats kan komen van de sturing door de sterk gereduceerde visuele kode van het notenschrift.